INVESTIGATION OF CANDIDATE IMPRINTED TUMOUR SUPPRESSOR GENES

Hannah Judson

Degree of Doctor of Philosophy

The University of Edinburgh

2001

DECLARATION

I declare that:

a) This thesis has been composed by myself.

b) The work is either my own, or the work/author involved is stated.

ACKNOWLEDGEMENTS

I would like to acknowledge the Medical Research Council for the studentship that enabled me to carry out this work, and would like to mention the following people for kindly providing tumour samples: Dr. Frank Speleman, Dr. Valerie Speirs, Dr. Amanda McCann, and Professor Ken Maclennan.

I would also like to thank my family for their encouragement, and my friends and colleagues in both Edinburgh and Leeds, for putting up with me through the stressful moments…(of which there were many...)

Finally, my particular thanks go to Professor David Bonthron for always having time to help me, and for all his guidance and support during the past three years

ABSTRACT

Imprinting is an epigenetic phenomenon that silences one allele of a gene, so that expression in one or more cell types is exclusively monoallelic, and dependent on parental origin. Approaches used to identify novel imprinted genes rely on characteristic features such as the clustering of imprinted genes, or their association with differentially methylated CpG islands.

An imprinted tumour suppressor gene involved in pathogenesis of neuroblastoma is believed to lie within chromosome 1p36. In this region, a search was initiated for imprinted genes in the vicinity of the imprinted gene *TP73*. The *DFFB* gene, encoding the apoptotic nuclease DNA fragmentation factor, was identified, and its intron-exon structure was elucidated. A pseudogene was also identified on chromosome 9. The tumour suppressor candidacy of *DFFB* was assessed through a comprehensive mutation screen of 42 neuroblastoma DNAs. No tumour-specific mutations were identified. Imprinting was then assessed by RT-PCR, which revealed biallelic expression of *DFFB*. It is unlikely that *DFFB* acts as a tumour suppressor in neuroblastoma.

During a systematic screen, a differentially methylated CpG island, NV149, had been identified. In the present study, this locus was mapped to 6q24. The nearest gene was found to be the cell-cycle control gene, *ZAC*. Monoallelic expression of *ZAC* from the paternal allele only was demonstrated in a range of fetal tissues. *ZAC* may possess a dual role in disease, such that upregulation by paternal duplication or paternal uniparental disomy of chromosome 6 results in transient neonatal diabetes mellitus (TNDM), whereas loss or down regulation of *ZAC* results in loss of cell cycle control, and hence tumorigenesis. Through analysis of a panel of B cell lymphomas, evidence was found for hypermethylation of the NV149 CpG island, which may be one mechanism through which expression of *ZAC* is lost in tumours.

ABBREVIATIONS

bp	Base pairs
BSA	Bovine serum albumin
CHEF	Contour-clamped homogeneous electric field
cR	Centiray
DEPC	Diethyl pyrocarbonate
DMR	Differentially methylated region
DMSO	Dimethyl sulfoxide
EDTA	Ethylenediaminetetraacetic acid
IPTG	Isopropylthio-β-D-galactoside
kb	Kilobase
LOH	Loss of heterozygosity
LOI	Loss of imprinting
Mb	Megabase
PAC	P1 artificial chromosome
PCR	Polymerase chain reaction
PNK	Polynucleotide kinase
psi	Pounds per square inch
rpm	Revolutions per minute
RT-PCR	Reverse transcriptase PCR
SAP	Shrimp alkaline phosphatase
SDS	Sodium dodecyl sulphate
SSC	Standard saline citrate buffer
TAE	Tris acetate electrophoresis buffer
TBE	Tris borate electrophoresis buffer
TCA	Trichloroacetic acid
TEMED	N,N,N'N'-tetramethylenediamine
UPD	Uniparental disomy
utr	Untranslated region
YAC	Yeast artificial chromosome

CONTENTS

TABLE OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 IMPRINTING

Genomic imprinting is an intriguing epigenetic phenomenon. The defining characteristic of imprinted genes is monoallelic expression, which defies the underlying principles of traditional Mendelian inheritance. Until the discovery of imprinting, the concepts of biparental inheritance and expression had remained unchallenged, and it was widely accepted that the maternal and paternal alleles of a gene were functionally equivalent, and that both were required for normal development.

Differential expression of the alleles of an imprinted gene is determined by their parental origin, such that in the case of a maternally imprinted gene, the paternal copy is expressed whereas the maternal copy is silenced. Conversely, a paternally imprinted gene will be expressed solely from the maternally inherited allele.

1.1.1 The discovery of imprinting

Parent-of origin effects were described in the two rodent species *Peromyscus polionotus* and *Peromyscus maniculatus*, in which a cross between a female of the former species and a male of the latter resulted in offspring that were 40% smaller than either parent, whereas the reciprocal cross produced oversized offspring that often died before birth (Dawson, 1965). The first formal demonstration of imprinting, however, was made through a series of experiments that demonstrated the non-equivalence of the parental genomes in the developing embryo (Barton *et al.* 1984; Surani *et al.* 1984; McGrath, Solter 1984). A key experiment was that of Surani *et al*. (Surani *et al.* 1984), in which reconstituted diploid eggs were generated using haploid parthenogenetic eggs as recipients for either a paternal or a maternal pronucleus derived from fertilized eggs. The critical observation was that those receiving a male pronucleus sometimes developed to term, whereas those receiving a female pronucleus generally showed poor development. Specifically, the few eggs in this second category that did manage to implant were notably retarded in growth, and

had very poorly developed extraembryonic tissues and trophoblast. The conclusions that were drawn from these results were that both male and female pronuclei are essential to allow full term development. A similar conclusion regarding the non-equivalence of the maternal and paternal genomes was reached by Cattanach *et al.* (Cattanach, Kirk 1985), who, taking a different approach, generated mice with uniparental disomies for particular chromosomal regions. In several cases, the observed phenotypic effects were dependent on the parental origin of the disomy, indicating that there is a differential contribution from the maternal and paternal genomes.

1.1.2 Imprinting mechanisms

Gaining an understanding of the mechanisms that govern the establishment and maintenance of the imprint is a key area of imprinting research (Constancia *et al.* 1998; Pfeifer 2000; Reik, Walter 1998). What is the imprint mark? How and when is it established in the germline, and how is it maintained in somatic cells? Definitive answers to these questions have not yet emerged but elucidating the mechanisms that underlie the possible erasure, establishment and reprogramming of the imprint will be a significant advance in this field of research.

1.1.2.1 Nature of the imprint

In taking a logical view of the process of imprinting, it is clear that the parental origin of chromosomes must be distinguished by a mark of some kind, which directly or indirectly gives rise to the monoallelic expression that we recognise as the defining feature of imprinting. The imprint, which in effect 'labels' each chromosome as either maternal or paternal, must possess several properties. Firstly, it must have stability, since it must be maintained throughout cell division and differentiation. Secondly, it must be consistently recognisable, since it is 'read' in the somatic cells, and converted into differential expression. Thirdly it must be reversible, since the parental nature of the imprint may require to be switched between successive generations. This reprogramming of the imprint is crucial,

had very poorly developed extraembryonic tissues and trophoblast. The conclusions that were drawn from these results were that both male and female pronuclei are essential to allow full term development. A similar conclusion regarding the non-equivalence of the maternal and paternal genomes was reached by Cattanach *et al.* (Cattanach, Kirk 1985), who, taking a different approach, generated mice with uniparental disomies for particular chromosomal regions. In several cases, the observed phenotypic effects were dependent on the parental origin of the disomy, indicating that there is a differential contribution from the maternal and paternal genomes.

1.1.2 Imprinting mechanisms

Gaining an understanding of the mechanisms that govern the establishment and maintenance of the imprint is a key area of imprinting research (Constancia *et al.* 1998; Pfeifer 2000; Reik, Walter 1998). What is the imprint mark? How and when is it established in the germline, and how is it maintained in somatic cells? Definitive answers to these questions have not yet emerged but elucidating the mechanisms that underlie the possible erasure, establishment and reprogramming of the imprint will be a significant advance in this field of research.

1.1.2.1 Nature of the imprint

In taking a logical view of the process of imprinting, it is clear that the parental origin of chromosomes must be distinguished by a mark of some kind, which directly or indirectly gives rise to the monoallelic expression that we recognise as the defining feature of imprinting. The imprint, which in effect 'labels' each chromosome as either maternal or paternal, must possess several properties. Firstly, it must have stability, since it must be maintained throughout cell division and differentiation. Secondly, it must be consistently recognisable, since it is 'read' in the somatic cells, and converted into differential expression. Thirdly it must be reversible, since the parental nature of the imprint may require to be switched between successive generations. This reprogramming of the imprint is crucial,

because an allele of paternal origin in one generation, for example, may subsequently reside in the female germline, at which point the imprint must be reset or switched to represent a maternal origin.

It is now well established that differential, parent-specific methylation of nearby CpG islands is a hallmark of imprinted genes, and it seems likely that methylation plays a critical role in either establishing or maintaining the imprint mark. As a candidate medium for carrying the information it satisfies all of the criteria mentioned above, in that it is stable, recognisable by the transcriptional machinery, and reversible. Differential methylation was first demonstrated for the genes *H19* (Ferguson-Smith *et al.* 1993; Bartolomei *et al.* 1993), *Igf2* (Sasaki *et al.* 1992), and *Igf2r* (Brandeis *et al.* 1993), and in agreement with a mechanistic function of this differential methylation, a loss of imprinting of all three genes was observed in mice that were deficient in DNA methyltransferase (*Dnmt1*) activity (Li *et al.* 1993).

1.1.2.2 Imprinting in development

Given the necessary properties of the imprint that have been described above, it is predicted that an imprint that signals the parental origin of the chromosomes should be resistant to the global genomic demethylation that occurs in pre-implantation development (Monk *et al.* 1987). However, many imprinted genes with a differentially methylated region do appear to become demethylated at this time, along with other non-imprinted genes (Brandeis *et al.* 1993). The parental identity of the chromosomes in the embryo must still be recognised though, in order for the correct imprints to be maintained and manifested in the form of monoallelic expression in the somatic cells. This implies that not all differentially methylated regions (DMRs) act as primary signals of parental origin. Some studies have now identified specific regions of differential methylation that are inherited from the gametes, and retained through all stages of development, including the global demethylation, as predicted. This suggests that there is a subset of differentially methylated regions that act as primary gametic imprints, stably marking the chromosomes with their origin. To date, at least six such putative primary gametic

because an allele of paternal origin in one generation, for example, may subsequently reside in the female germline, at which point the imprint must be reset or switched to represent a maternal origin.

It is now well established that differential, parent-specific methylation of nearby CpG islands is a hallmark of imprinted genes, and it seems likely that methylation plays a critical role in either establishing or maintaining the imprint mark. As a candidate medium for carrying the information it satisfies all of the criteria mentioned above, in that it is stable, recognisable by the transcriptional machinery, and reversible. Differential methylation was first demonstrated for the genes *H19* (Ferguson-Smith *et al.* 1993; Bartolomei *et al.* 1993), *Igf2* (Sasaki *et al.* 1992), and *Igf2r* (Brandeis *et al.* 1993), and in agreement with a mechanistic function of this differential methylation, a loss of imprinting of all three genes was observed in mice that were deficient in DNA methyltransferase (*Dnmt1*) activity (Li *et al.* 1993).

1.1.2.2 Imprinting in development

Given the necessary properties of the imprint that have been described above, it is predicted that an imprint that signals the parental origin of the chromosomes should be resistant to the global genomic demethylation that occurs in pre-implantation development (Monk *et al.* 1987). However, many imprinted genes with a differentially methylated region do appear to become demethylated at this time, along with other non-imprinted genes (Brandeis *et al.* 1993). The parental identity of the chromosomes in the embryo must still be recognised though, in order for the correct imprints to be maintained and manifested in the form of monoallelic expression in the somatic cells. This implies that not all differentially methylated regions (DMRs) act as primary signals of parental origin. Some studies have now identified specific regions of differential methylation that are inherited from the gametes, and retained through all stages of development, including the global demethylation, as predicted. This suggests that there is a subset of differentially methylated regions that act as primary gametic imprints, stably marking the chromosomes with their origin. To date, at least six such putative primary gametic

imprints have been detected, associated with the genes *H19*, *Igf2r*, *Gnas*, *U2afbp-rs/U2af1-rs1*, *Snrpn*, and *Kcnq1*. The best studied of these is the gametic imprint mark associated with the maternally expressed imprinted gene *H19*.

A bisulphite analysis of the paternal specific *H19* methylation during spermatogenesis has shown that a 2kb region of paternal-specific methylation is inherited from sperm, and stably propagated and maintained in the preimplantation embryo (Tremblay *et al.* 1997; Tremblay *et al.* 1995). It is believed that this 2kb DMR subsequently extends, to render the *H19* promoter and gene body of the paternal allele hypermethylated (Pfeifer 2000). This 2kb paternal-specific methylation upstream of *H19* appears to represent a primary, gametic imprint since it is present in sperm, absent from oocytes, and is conserved at all stages of embryogenesis that have been analysed.

Other than *H19*, potential gametic imprint marks have been reported for at least five other genes: *Gnas* apparently has a region of differential methylation that is present in oocytes, but not in spermatozoa, and is maintained at preimplantation stages of development (Liu *et al.* 2000a); *U2afbp-rs/U2af1-rs1* carries an oocyte-specific methylation imprint mark that is maintained throughout development, and like the *H19* paternal specific imprint mark, the methylation is expanded before implantation to assume the final methylation pattern (Shibata *et al.* 1997); *Igf2r* has also been shown to carry a gamete-specific methylation mark that is preserved during early embryogenesis (Brandeis *et al.* 1993). The mouse *Snrpn* gene has two regions of differential methylation that are inherited from the gametes, and are protected from methylation changes in the embryo, and thus represent a gametic imprint mark (Shemer *et al.* 1997). Similarly, the CpG island of the *Kcnq1* gene associated with the antisense transcript *Kcnq1ot1*, is methylated in oocytes, unmethylated in sperm, and these differences appear to be stably conserved in the embryo (Engemann *et al.* 2000).

The discovery of primary gametic imprint marks has indicated that imprinting may be under the control of a small number of more permanent methylation signals,

which direct the establishment of other DMRs associated with imprinted genes within the locus. If this is the case, then it stands to reason that each imprinted domain of the genome may be dominated by such a signal, and that more are yet to be identified.

1.1.2.3 Imprint erasure and reprogramming in the germline

Much of the investigation in this field has attempted to address the questions of how and when the imprint can be erased from the chromosomes and reset in either the same or opposite parental pattern for subsequent generations. One school of thought also questions whether it is actually necessary for the imprint to be erased from the parental chromosome that is of the same sex of the germline – for example, in the male germline does the imprint derived from the sperm become erased only to be reset with the same paternal mark in the germline, or are same-sex imprints retained?

Irrespective of which scenario is accurate, it is widely believed that the reprogamming of imprints must occur during gametogenesis when the parental genomes are entirely independent from each other. The gametes in both males and females originate from the primordial germ cells (PGCs), which migrate from the epiblast, through extraembryonic tissue, to the gonad primordia in the genital ridge. In the mouse they reach their destination by 10.5 to 11.5 days postcoitum (dpc). Until this point their development is identical in both male and female embryos, but subsequent to this, in the female germline, the cells enter meiotic prophase, and oocyte growth proceeds after birth. In the male germline, the germ cells undergo mitotic arrest until after birth when they enter mitosis again, and subsequently undergo meiotic differentiation (Constancia *et al.* 1998).

The majority of studies have analysed the establishment of the imprints in the mouse, and nearly all of the work has examined the *H19* imprint, since this is one of the best characterised imprinted genes. One of the key points in the reprogramming of the germline imprints is a demethylation event in the primordial germ cells early in fetal development, which is followed later by a *de novo* methylation event (Razin, Shemer

1995). Estimates of the exact timing of this process are variable, but suggest that the demethylation occurs between 8.5dpc and 12.5dpc. Accurate estimates are difficult due to the technical problems associated with analysing material of early stages such as these. The finding of biallelic expression of *H19* in PGCs of the genital ridge at 11.5dpc led Szabo and Mann (Szabo, Mann 1995) to infer that imprinting is already erased, or at least not recognised in the genital ridge by this stage. However, observation of such biallelic expression does not necessarily prove that imprints are completely erased at this stage – as described, *H19* has a gametic imprint in the form of the upstream 2kb DMR, but this may not be directly responsible for allelic silencing of the paternal allele. Rather, it may be responsible for marking the parental origin of that chromosome, and acting as a signal for the further methylation that colonises the downstream promoter and gene at a later stage. So, biallelic expression at this early stage, and paternal-specific methylation of the 2kb DMR are not necessarily mutually exclusive.

Ueda *et al.* (Ueda *et al.* 2000) attempted to determine the timing of the *H19* imprint reprogramming in mouse. They showed that in both male and female PGCs, the *H19* imprint was incompletely erased at 12.5 and 13.5dpc despite the global demethylation. In the maternal germline, the *H19* imprint of the paternally derived chromosome was completely erased on entry into meiosis at 16.5dpc, whereas in the paternal germline, the partial methylation was overlaid with full methylation as a continuous process in mitotically arrested gonocytes of fetal and postnatal development. Interestingly, Ueda *et al.* noted that the establishment of the new imprint in the paternal germline occurred more quickly on the paternal chromosomes than on the maternal chromosomes. Overall, it could be interpreted from the results of this study that the methylation imprint of *H19* is not completely erased, but is maintained in at least some PGCs. However, it is not possible to conclude that this is definitely the case, since a complete erasure of the imprint may have occurred at earlier stages than were investigated.

Kerjean *et al.* (Kerjean *et al.* 2000) performed one of the few studies of imprint establishment in human. They analysed *H19* and the oppositely imprinted,

maternally methylated gene *PEG1/MEST*. Their main finding, in contrast to that of Ueda *et al.* (Ueda *et al.* 2000), was that the original methylation imprints of both genes were completely erased before the fetal spermatogonia stage in the male germline. Subsequent to this, and in agreement with Ueda *et al.*, they showed that the *H19* imprint is re-established in spermatogonia. Conversely, *PEG1/MEST* remains unmethylated in the male germline.

Davis *et al.* (Davis *et al.* 1999) discovered that in postnatal male germ cells the mouse *H19* DMR is methylated on the paternal chromosome, and also on a significant proportion of the maternal chromosomes. As with the results of Ueda *et al.*, this implies differential acquisition of the paternal imprint on the chromosomes in the male germline during spermatogenesis. A more recent study by Davis *et al.* (Davis *et al.* 2000) has reiterated this point, that the paternal allele acquires the *H19* DMR methylation before the maternal allele. However, Davis *et al.* make the critical point that if methylation imprints are completely erased before this point (as they have shown them to be), then the maternal and paternal chromosomes remain distinguishable and non-equivalent even in the absence of methylation. There must therefore be some mark, other than methylation, which continues to signify the parental origin of the chromosomes, even though it is not necessary for them to be distinguishable at this point. Davis *et al.* (Davis *et al.* 2000) propose a model to explain this. They point out that since the paternal allele appears to acquire its methylation at a similar timescale to other genes, between 15.5 and 18.5dpc (Kafri *et al.* 1992), it is the maternal allele that is somehow resistant or less receptive to methylation. As discussed further in 1.2.1.1.3, the maternal allele of *H19* is bound by the CTCF protein (CCCTC binding factor) due to the unmethylated status of the DMR on this allele. This functions as a boundary element in the regulation of reciprocal imprinting of the *H19* and *IGF2* genes. However, when carried by the maternal chromosome into the male germline, this additional epigenetic marking may initially prevent access of methyltransferases. Later in development, the binding of CTCF to the maternally derived chromosome may be lost by degradation of the protein or other means, resulting in methylation of the maternal allele so that it assumes the correct paternal identity in the sperm. Thus, a link is made between the

role of methylation, and the structure of chromatin, in marking the parental identity of the chromosomes, and in establishing the imprint in the germline.

1.1.2.4 Other possible epigenetic mechanisms

Methylation has been a focus in the attempts to understand the mechanisms of imprinting. However, as suggested by the association of differential methylation with binding of the CTCF protein in the *H19* DMR (a full explanation of this is given later in 1.2.1.1.3), the role of methylation may be to determine the chromatin structure. Thus, chromatin structure may be the element that controls the monoallelic expression, whereas methylation may be the underlying imprint that determines the establishment of local chromatin structure.

Analysis of chromatin structure has utilised the DNaseI footprinting techniques. Hark and Tilghman (Hark, Tilghman 1998) identified hypersensitive sites that were specific to the unmethylated maternal *H19* allele. Similarly, Feil *et al.* (Feil *et al.* 1997) detected hypersensitive sites on the unmethylated paternal allele of the *U2af1-rs1* gene, but not on the methylated maternal allele. Both studies suggest that the differential methylation of the two parental alleles may act to allow or prevent binding of non-histone proteins, which determine the chromatin organisation. Along these lines, Watanabe *et al.* (Watanabe *et al.* 2000) correlated imprinting with differential chromatin packaging. They showed that the chromatin of the non-transcribed allele of *Igf2r* was packaged more compactly than in the transcribed allele, whereas this difference was not found in the case of a nearby non-imprinted gene, *Sod2*.

Changes in chromatin packaging may determine which alleles are transcribed. Additionally, it has been suggested that the binding of proteins to DMRs, and the specialised chromatin structure may protect the unmethylated allele of imprinted genes against *de novo* methylation, in a role of maintenance (Bird 1992). It might also be expected therefore, that the methylated allele might be protected against demethylation. Some studies have indicated that proteins that bind exclusively to

methylated DNA, e.g. MeCPs, may be involved. Indeed, the protein MeCP2 has been demonstrated to bind to methylated DNA, and recruit a multi-protein complex that includes histone deacetylases, which could have the effect of decreasing the levels of histone acetylation, thus altering the conformation of the chromatin towards a hypoacetylated, transcriptionally inaccessible state (Jones *et al.* 1998; Nan *et al.* 1998).

The emerging picture seems to indicate, therefore, that the imprinting mechanism may be a process that encompasses not only methylation, but also alterations in chromatin structure, directed by the actions of specific proteins.

1.1.3 Evolution of imprinting

It is clear that imprinting is required for normal development. The extent to which aberrant, defective or lost imprinting can cause disease will be discussed later in this chapter. However, the question remains as to why imprinting evolved initially, and why it is evolutionarily conserved in mammals and also in some plants. There is a significant cost to those organisms that imprint parts of their genome, because the protection against deleterious mutations granted by diploidy is lost. Imprinting renders the genome partly hemizygous, and since nearly all imprinted genes are known to have critical roles in development, the evolutionary benefits of imprinting must therefore outweigh the disadvantages of the vulnerability that is associated with hemizygosity. There must be a reason why mammals imprint a subset of their genes, and why this subset appears to include those with a critical developmental role, thus conveying an increased susceptibility to disease, particularly cancer.

1.1.3.1 Hypotheses to explain the evolution of imprinting

To date, there have been several proposed hypotheses attempting to offer an explanation for the evolution of imprinting. These have included ideas suggesting that imprinting acts effectively as a speciation mechanism in mammals (Varmuza 1993), as a host defence system to protect against invasion by the DNA of foreign

organisms (Barlow 1993), or as a surveillance mechanism allowing detection of chromosome loss (Thomas 1995). However, perhaps the two most convincing proposals are that imprinting arose as a means to protect females against trophoblastic disease (Varmuza, Mann 1994), or that it exists as a result of the different selection pressures, and thus competition, between the maternal and paternal genomes within the developing embryo (Moore, Haig 1991). These two hypotheses are discussed below.

1.1.3.2 Imprinting as a mechanism of protection against trophoblastic disease

One of the more feasible models proposes that imprinting may serve to limit the potential malignancy of ovarian germ cell tumours. It has been estimated that 25% of benign ovarian tumours that arise in humans are derived from parthenogenetically-activated oocytes (Varmuza, Mann 1994). These oocytes develop *in situ*, but have particularly undeveloped extra-embryonic tissues. The significance of this is that the trophoblast, by definition, has a role of invasion and vascularisation, and in the absence of imprinting of the genes responsible for trophoblast development, these tumours might possibly be more invasive, metastatic, and consequently more lethal. Varmuza and Mann (Varmuza, Mann 1994), propose that the selective pressure for maternal imprinting of these genes is therefore high, and offers protection to females against trophoblastic disease. The model has however attracted strong criticism, and has been challenged on the basis of several weak points (Solter 1994; Haig 1994; Moore 1994). Firstly, it does not offer a convincing evolutionary explanation for the imprinting of paternally derived genes, because it accounts only for the imprinting of genes that would act to restrain the development of the trophoblast, and suggests that only the maternal genome has a role in imprinting. Similarly, it does not rationalise the imprinting of genes that have any other function unrelated to trophoblast development. The authors account for both of these points by suggesting that some genes become imprinted through being 'innocent bystanders', rather than requiring to be imprinted for a valid functional reason. This idea has not been generally accepted, because there is no explanation for the evolutionary maintenance of

imprinting in these genes. Additionally, it does not account for the existence of imprinting in any plants, yet imprinting has been observed in plant genes (Chaudhuri, Messing 1994). This model, whilst offering a novel idea to explain the evolution of imprinting, is perhaps not sufficiently robust to account for our current understanding of imprinting.

1.1.3.3 The parental conflict hypothesis

The most widely accepted hypothesis for the evolution of imprinting is the parental conflict model (Moore, Haig 1991). This model identifies the different selection pressures that operate on paternally and maternally derived alleles in an embryo, suggesting that they have conflicting interests, and are therefore in competition with each other. The model applies to any species in which multiple paternity is a possibility, because it is in the interests of the maternal genome to limit fetal growth in order to restrain the demands made on the maternal resources, such that the success of future pregnancies is protected. The paternal genome however, does not share these interests, and is not subject to selection pressures that benefit other fetuses. The paternal interests in a fetus are therefore selfish, to the cost of both the mother and future, paternally-unrelated offspring. The nature of the model therefore makes several predictions. Of greatest importance is that maternally expressed genes will be growth suppressing, and paternally expressed genes will be growth enhancing. The second prediction is that imprinting should be crucial in embryo development, but its effects are not expected to persist after weaning. Thirdly, the degree to which multiple paternity occurs in a species should be reflected in the 'intensity' of imprinting (Hurst, McVean 1998), although evolutionary rates could affect this, and it may not be valid to make a judgement of the intensity of imprinting given that some species are evolving more rapidly than others.

For many years the parental conflict model has been accepted as the most likely explanation for the evolution of imprinting. Early discoveries of imprinted genes were felt generally to support the model. In particular, in the mouse, the genes *Igf2* and *Igf2r* represent prime examples conforming to the conflict model. The paternally

expressed *Igf2* gene encodes a growth factor, and functions to promote growth, whereas the oppositely imprinted maternally expressed gene *Igf2r* encodes a receptor for Igf2, and in effect is a suppressor of growth because it activates growth inhibitors, and targets Igf2 for lysosomal degradation (Hankins *et al.* 1996). In accordance with the parental conflict model, the functions of these reciprocally imprinted genes are antagonistic, or perhaps mutually interactive (Hurst, McVean 1998). Additionally, the respective growth enhancing and growth suppressing properties of these genes are notably lost in a parent-specific way in transgenic knockout mice, with growth deficiency occurring in paternally *Igf2*-deleted mice, and fetal overgrowth occurring in the absence of maternal *Igf2r* (DeChiara *et al.* 1991; Lau *et al.* 1994). However, it is now known that the equivalent human gene, *IGF2R*, is not imprinted. The conflict model has recently been critically reviewed (Hurst, McVean 1998), and the flaws are beginning to emerge. An increasing number of exceptions to the model are becoming apparent, in terms of genes that do not display the predicted effects in knockout experiments, such as *Snrpn*, which is paternally expressed, but shows no phenotypic consequences when an intragenic region is knocked out (Yang *et al.* 1998); and *ZAC*, which is the subject of Chapters 4 and 5 of this thesis, and is a growth suppressor despite being paternally expressed. Post-weaning effects that contravene the predictions of the model are seen in the case of the paternally expressed gene *Peg1/Mest*. Lefevbre *et al.* (Lefebvre *et al.* 1998) generated a disruption of the *Mest* gene in embryonic stem cells, and found that when transmitted paternally, the resulting *Mest*-deficient female offspring showed abnormal maternal behaviour in the care of their litter, including absence of placentophagia. This phenotype was the first demonstrated imprinting effect in adults, but the imprinting of a gene that affects adult behaviour is not predicted by the conflict model. Furthermore, the discovery of imprinting in the obligately selfing plant *Arabidopsis thaliana* is neither predicted nor accounted for by the model, because the reproductive nature of this species presents no selective difference between the paternal and maternal genomes (Hurst, McVean 1998).

In summary, it would appear that whilst the predictions of the parental conflict model are fulfilled by the features of many imprinted genes, it cannot fully explain the

evolution of imprinting since there are several exceptions to the model. As suggested by Hurst and McVean (Hurst, McVean 1998), the acceptance of this model has been more due to the lack of a convincing alternative, than to the compelling nature of the model itself.

1.2 IMPRINTED GENES

There are now at least thirty known imprinted human genes, but estimates of the total number of imprinted genes suggest that the number could be as high as one or two hundred (Barlow 1995). Most of the known imprinted genes have been identified on the basis of the classic, diagnostic features of imprinting, namely monoallelic, parent-of-origin expression, and association with differentially methylated CpG-rich regions. A representative selection of imprinted genes is discussed below, in relation to some of the characteristic traits of imprinted genes, such as clustering in the genome, the involvement of regulatory imprinting centres, and the possible regulation of imprinting by antisense transcripts, in addition to some of the more unusual features, such as isoform-, promoter- and tissue-specific imprinting.

1.2.1 Clustering of imprinted genes

The clustering of imprinted genes has become recognised as a classic feature of imprinting, and as more genes are identified, the extent to which clustering occurs will become apparent. Currently, the majority of known imprinted genes are part of the two major clusters, at human chromosome 11p15.5, and 15q11-q13. However, there is at least one potential new cluster emerging from recent research on chromosome 7. The clustering of imprinted genes raises obvious questions about regional control and coordinate regulation, and about the evolution of such regions. The 11p15.5 cluster is discussed below.

1.2.1.1 Cluster at 11p15.5

Perhaps the most intensively studied imprinted genes to date are those which reside in a cluster on human chromosome 11p15.5. This cluster, spanning a megabase, contains the first identified imprinted gene *IGF2*, along with at least ten other genes, of which the majority are imprinted. *IGF2* itself has been well characterised, yet a complete picture of the regulation and maintenance of imprinting in this region remains elusive.

1.2.1.1.1 *Igf2*, the first known imprinted gene

Igf2 encodes a growth factor with a crucial role in embryonic growth. It is expressed in many tissues at high levels during development, and is thought to function in an autocrine/paracrine manner (DeChiara *et al.* 1991). *Igf2* was the first known imprinted gene in mouse; imprinting was demonstrated through a series of transgenic experiments in which the major observation was that transmission of a null *Igf2* mutation through the paternal germline resulted in growth deficient heterozygous progeny, whereas transmission of the same mutation through the maternal germline resulted in phenotypically normal mice (DeChiara *et al.* 1991). This result indicates that only the paternally inherited copy of *Igf2* is functionally active. Additional evidence for parental imprinting of *Igf2* from this study was apparent in the fact that homozygous mutant mice (carrying the deletion on both chromosomes) and heterozygous *Igf2* mutants (in which the deletion had been inherited paternally) were phenotypically indistinguishable. DeChiara *et al.* also demonstrated switching of the imprint through the germline in successive generations by mating each type of heterozygote (i.e. those with the paternally-derived deletion, and those with the maternally-derived deletion) with wild type siblings. Given that allele switching is a key requirement of imprinting, these experiments provided indisputable evidence for the imprinting of *Igf2*.

1.2.1.1.2 *IGF2* and *H19* are reciprocally imprinted

In humans, *IGF2* is closely linked to the *H19* gene on chromosome 11p15.5. *H19* itself encodes an untranslated RNA. The suggestion that the genes may be coordinately regulated came from the observations that they lie in close proximity to each other, and are reciprocally imprinted – *IGF2* is paternally expressed, and maternally silenced, whilst *H19* is maternally expressed and paternally silenced (Mutter *et al.* 1993).

1.2.1.1.3 Proposed mechanisms for imprinting regulation of *IGF2* and *H19*

The characteristic clustering of imprinted genes implies that regulation of imprinting may, in some cases, be controlled by common, regionally acting elements. In the case of *IGF2* and *H19*, which are reciprocally imprinted, a model was proposed which suggests that these two genes compete for the action of the enhancers which lie downstream of *H19* (Leighton *et al.* 1996). By the enhancer competition model, the *H19* gene on the maternal chromosome has exclusive use of these enhancers, either because of its closer proximity than *IGF2*, or because of the *H19* promoter having some advantage in enhancer preference over *IGF2*. Either way, the maternal *H19* gene monopolises the enhancers, and the outcompeted maternal *IGF2* gene remains silent. On the paternal chromosome however, the allele-specific methylation of *H19* silences this copy, such that the competition from *H19* is eliminated, and the upstream gene *IGF2* is therefore granted exclusive use of the enhancers (Schmidt *et al.* 1999).

The enhancer competition model was widely accepted as a potential mechanism for regulating the imprinting of *IGF2* and *H19*, and several experiments have provided substantiating evidence (Leighton *et al.* 1995). However, more recently, experimental evidence has failed to fulfill the predictions of the enhancer competition model, and has brought to light the possibility that imprinting regulation may occur by an alternative mechanism. The evidence against the enhancer competition model lies in the results of experiments by Schmidt *et al.* (Schmidt *et al.* 1999), in which they generated a deletion of the *H19* promoter. The enhancer

competition model predicts that the outcome of the resulting loss of *H19* transcription (and elimination of the competition) would cause transcription of the previously silent maternal *IGF2* allele. This prediction was not fulfilled, as the imprinting of the *IGF2* gene was not affected by the deletion. If transcription from the *H19* promoter is not the factor that prevents transcription of *IGF2*, then another mechanism of imprinting regulation must be involved (Schmidt *et al.* 1999).

Recent research has established an alternative model, which is becoming increasingly favoured over the original enhancer competition model. The model proposes that 'insulator' elements located in the imprinting control region between *H19* and *IGF2*, act to delimit the boundaries between transcriptionally active euchromatin, and inactive heterochromatin (Wolffe 2000). The central concept of the model is that the DNA binding protein CTCF recognises and binds to the 21bp CpG-rich insulator sequences via a domain comprising 11 zinc fingers, to prevent activation of upstream *IGF2* by the enhancers located downstream of *H19*. The most crucial aspect of the model is that the binding of CTCF is dependent on the sequence being unmethylated. Thus, on the maternal chromosome, which is unmethylated at this region, CTCF binds, and insulates *IGF2*, preventing enhancer access, and effectively silencing the gene, whilst *H19* remains active. Conversely, on the paternal chromosome, which is methylated at this region, CTCF is unable to bind, so *IGF2* has full access to the downstream enhancers. The methylation on the paternal chromosome extends to encompass the *H19* promoter region and part of the gene, and this directly represses transcription. Through DNA footprinting experiments, Szabo *et al.* (Szabo *et al.* 2000) demonstrated the methylation-sensitive nature of CTCF binding *in vivo*; DNA footprints representing the presence of bound protein were identified on the unmethylated maternal chromosome at all four 21bp repeat sequences in the imprinting control region between *H19* and *IGF2*, but were not detected on the paternal chromosome. Similarly, Bell and Felsenfeld (Bell, Felsenfeld 2000) and Hark *et al.* (Hark *et al.* 2000) found evidence *in vitro* that CTCF was bound to all four insulator sites on the maternal chromosome, but when the chromosome was methylated prior to the assay, no CTCF binding was detected. Using gel shift assays and chromatin immunopurification analyses, Kanduri *et al.* (Kanduri *et al.* 2000)

contributed additional evidence to confirm that binding of the CTCF is methylation-sensitive and specific to the parental origin of the chromosome. The evidence accumulating from all of these studies is convincing, and indicates that CpG methylation in this region (as found on the paternal chromosome) can directly inhibit the binding of the CTCF protein. Interestingly, an earlier study (Thorvaldsen *et al.* 1998) had shown that maternal inheritance of a deletion of this region resulted in activation of the *IGF2* gene and downregulation of *H19* expression; an explanation for this is now apparent on the basis of these more recent findings, in that the deletion likely eliminated the insulator, preventing CTCF binding, and permitting activation of *IGF2* by the enhancers.

The precise action of the insulator, and its association with CTCF is still under scrutiny, but possibilities are that the insulator element determines the boundaries between transcriptionally active euchromatin, and inactive heterochromatin, or that it retargets or binds the transcription factors that are required for gene regulation (Wolffe 2000). The model itself may be more complex still, in that newly described tissue-specific 'silencer' elements may have a role in the repression of *IGF2* and *H19*, depending on the parental origin of the chromosome on which they reside (Ferguson-Smith 2000). Drewell *et al.* (Drewell *et al.* 2000) showed that deletion of a silencer element upstream of *H19* on the paternal chromosome resulted in activation of *H19*, without any effect on the insulator element or the expression of the *IGF2* gene. Clearly, our understanding of the regulation of this region is far from complete, but is being deciphered piece by piece. Figure 1 illustrates the enhancer competition model and the insulator model.

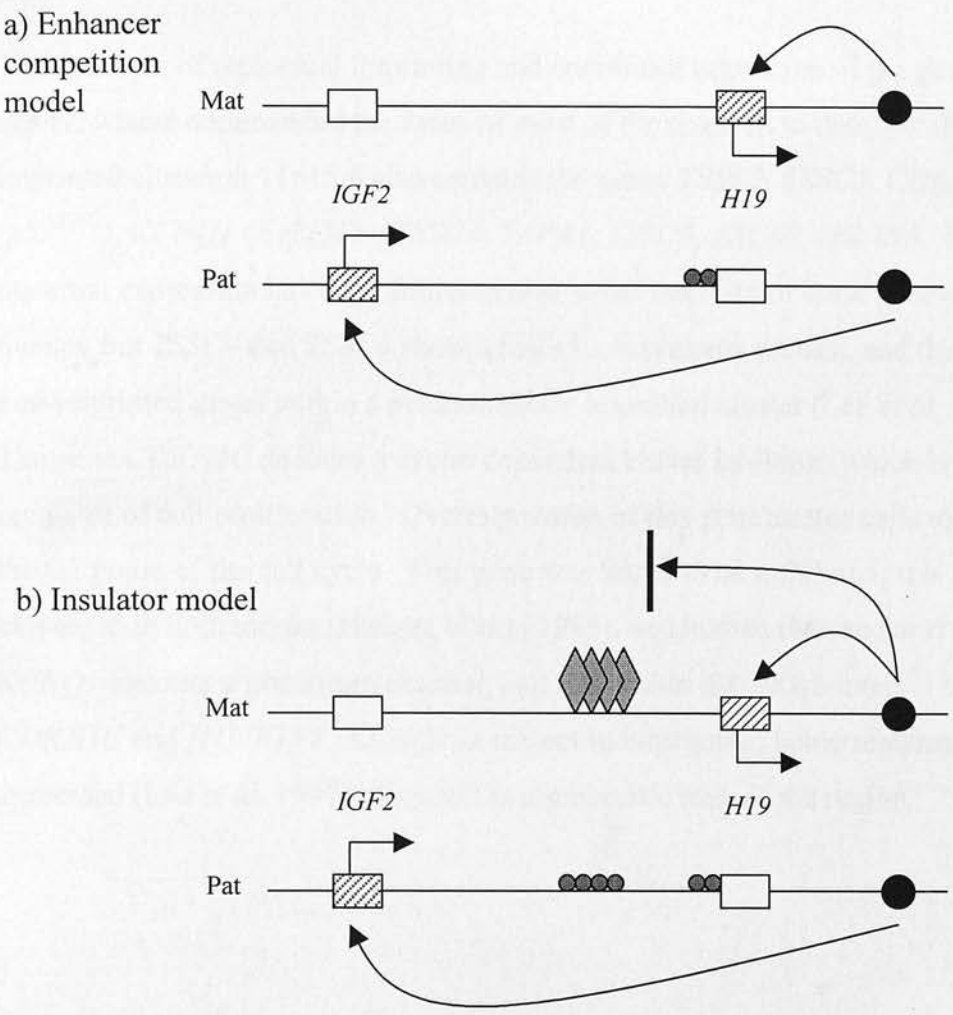


Figure 1. Proposed models for the regulation of imprinting at *H19/IGF2*. a) The enhancer competition model. *H19* and *IGF2* compete for the enhancers (large black circles) downstream of *H19*. Methylation (small grey circles) on the paternal chromosome prevents transcription of *H19*, allowing *IGF2* to use the enhancers. The maternal chromosome is unmethylated, so *H19* has full use of the enhancers. The expressed genes are shown by shaded rectangles; silenced genes are shown by open rectangles.

b) The insulator model. The CTCF protein (grey diamonds) binds to the unmethylated maternal chromosome, blocking access of enhancers to *IGF2*, so only *H19* is expressed. On the paternal chromosome, CTCF is prevented from binding because of methylation of the insulator sequence, so the enhancers can activate *IGF2*. Transcription of *H19* from this chromosome is prevented by methylation of the promoter.

Adapted from Maher *et al.* (2000) and Szabo *et al.* (2000). Not to scale.

1.2.1.1.4 Other genes, both imprinted and non-imprinted within the 11p15.5 cluster

The concepts of reciprocal imprinting and communal regulation of the genes *IGF2* and *H19* have commanded the focus of most of the research to date, yet the imprinted cluster at 11p15.5 also contains the genes *TSSC3*, *TSSC5*, *CDKN1C* (*p57^{KIP2}*), *KCNQ1* (*KvLQT1*), *TSSC4*, *TAPA1*, *TSSC6*, *ASCL2*, and *INS*. Exclusively maternal expression has been demonstrated in the majority of these genes in the human, but *TSSC4* and *TSSC6* show a biallelic expression pattern, and thus represent non-imprinted genes within a predominantly imprinted cluster (Lee *et al.* 1999a). The gene *CDKN1C* encodes a cyclin dependent kinase inhibitor, which is a negative regulator of cell proliferation. Overexpression of this gene causes cells to arrest in the G1 phase of the cell cycle. This gene was found to be imprinted; it is maternally expressed in both mouse (Hatada, Mukai 1995), and human (Matsuoka *et al.* 1996). *KCNQ1* encodes a potassium channel, and lies within the 700kb interval between *CDKN1C* and *H19/IGF2*. *KCNQ1* is subject to imprinting, being maternally expressed (Lee *et al.* 1997). Figure 2 is a schematic map of the region.

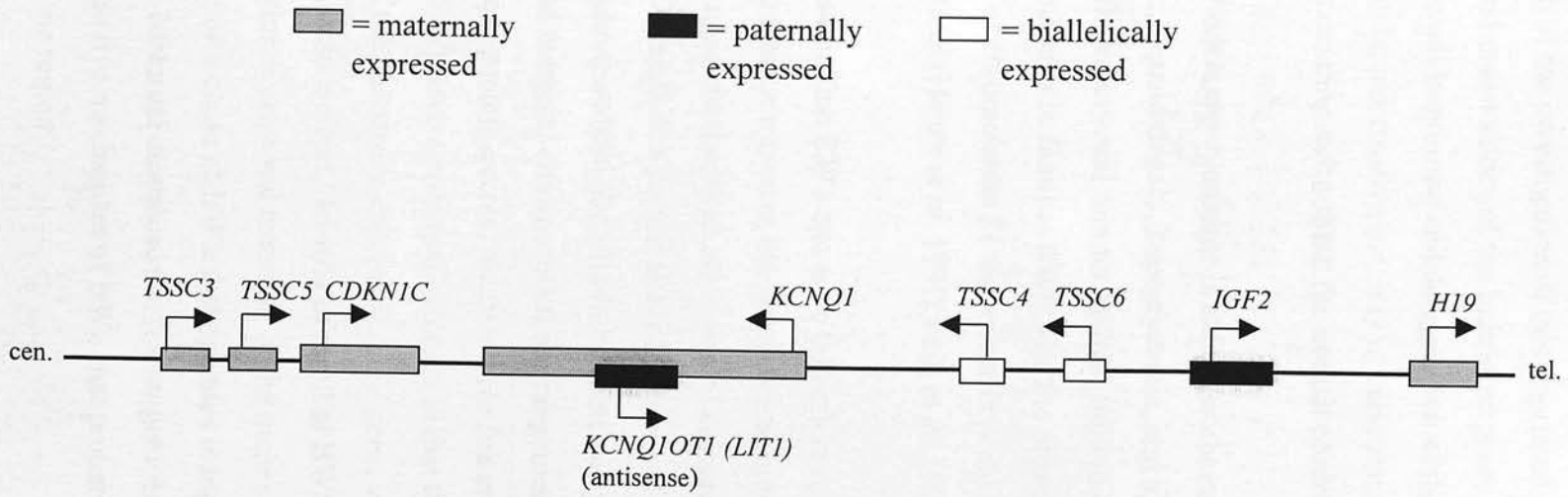


Figure 2. Schematic representation of the 11p15.5 imprinted gene cluster. The nature of expression of each gene is shown by the colour of the box, as indicated above. The arrows indicate the transcriptional orientation. Reproduced and adapted from Maher *et al.* (2000) and Engemann *et al.* (2000). Not to scale.

1.2.1.1.5 Beckwith-Wiedemann Syndrome (BWS) is associated with aberrant imprinting of one or more genes in this cluster

Whilst much of the investigation of this imprinted gene cluster has focussed on the regulation and maintenance of the imprinted genes under normal circumstances, the study of aberrant imprinting and deregulation of this cluster, and the consequential outcomes can be informative not only in understanding the resulting diseases, but also in understanding more about the normal control of imprinting.

Beckwith-Wiedemann syndrome is a disease characterised by prenatal overgrowth, macroglossia, organomegaly, hypoglycemia, and a predisposition to cancer, particularly the embryonal tumour Wilms' tumour (Reik, Maher 1997). The inheritance patterns in familial BWS, and the presence of uniparental disomy and duplications of chromosome 11 that occur in some cases, are suggestive of an imprinting effect (Henry *et al.* 1991; Reik *et al.* 1994).

It is now apparent that BWS can arise through several different mechanisms, some of which are alternative routes to the same genetic outcome, and some of which are unique. The main mechanisms are: paternal uniparental disomy for chromosome 11, paternal duplications of a region at 11p15.5, loss of imprinting at *IGF2* independent of structural chromosomal abnormalities, maternally-inherited mutations of *CDKN1C*, and maternal chromosomal rearrangements (Li *et al.* 1997; Reik, Maher 1997). Of these genetic events, the first three (paternal UPD11, paternal duplications and LOI at *IGF2*) have a common outcome in that they all lead to biallelic expression of the paternally-expressed *IGF2* gene, which could therefore have a significant causative effect. In 40% of familial BWS cases, the only identifiable genetic alteration is maternal mutation of the more centromeric gene *CDKN1C*, which represents a cause of BWS that operates independently of any *IGF2* involvement. Maternal chromosomal rearrangements are infrequent, but represent an additional causative mechanism of BWS that probably acts through disruption of imprinting in the region.

Loss of imprinting (LOI) of *IGF2*, and the resulting biallelic expression is the leading cause of sporadic BWS cases. It has been shown that LOI of *IGF2* can occur either in conjunction with *H19* hypermethylation, or conversely, completely independently of changes in *H19* methylation (Joyce *et al.* 1997). This discovery, in addition to the fact that imprinting of *KCNQ1* and *CDKN1C* is unaffected by deletions of *H19* (Caspary *et al.* 1998), is indicative of two separate, independent imprinting control centres in this cluster. A study by Smilinich *et al.* (Smilinich *et al.* 1999) has identified an antisense transcript, *KCNQ1OT1* (*LIT1*), which may be the key to the function of the second, *H19*-independent control centre. *KCNQ1OT1* is located in an intronic region of the *KCNQ1* gene, and is associated with a maternally methylated CpG island, and expressed exclusively from the paternal allele (Mitsuya *et al.* 1999).

In accordance with these ideas, it seems that BWS may be caused by mutations in either of these imprinting control regions. Mutations in the first control centre are associated with hypermethylation of the *H19* DMR, resulting in LOI and biallelic expression of *IGF2*. In these cases there is normal, maternal-specific methylation of the *KCNQ1OT1* CpG island. In contrast, cases in which the second putative imprinting control centre is involved display normal methylation patterns at *H19*, may have LOI and biallelic expression of *IGF2*, but are also accompanied by a loss of methylation at *KCNQ1OT1* that would be predicted to result in biallelic expression of this antisense transcript (Smilinich *et al.* 1999). It is clear from all of these results that there are two distinct imprinting centres. However, the second imprinting control centre, which is strongly associated with the demethylation of the *KCNQ1OT1* CpG island within *KCNQ1*, is less well understood than the *H19/IGF2* imprinting centre. In particular, there remains some confusion about whether the LOI at the *KCNQ1* CpG island, and the resulting biallelic *KCNQ1OT1* expression, are associated with LOI of downstream *IGF2*. Smilinich *et al.* had initially proposed the ideas regarding this imprinting control centre as an *H19*-independent mechanism of *IGF2* LOI because they identified two distinct groups of BWS patients: those with *H19* hypermethylation but normal *KCNQ1* imprinting; and those with normal *H19* imprinting, but LOI at the CpG island within *KCNQ1*. In both categories the patients

had biallelic *IGF2* expression. However, Lee *et al.* (Lee *et al.* 1999b) and Mitsuya *et al.* (Mitsuya *et al.* 1999) found LOI of *KCNQ1OT1* in BWS, but did not find an association with biallelic *IGF2* expression.

Important goals now include the elucidation of the mechanisms by which mutations in each of these imprinting centres can cause BWS, especially in the case of the putative *KCNQ1*-associated centre, which on present evidence may cause BWS independently of biallelic *IGF2* expression.

1.2.1.1.6 *KCNQ1* carries a gametic imprint mark

The evidence that has emerged from the study of BWS patients has suggested that two independent imprinting centres control the imprinting of genes in this cluster. The first of these domains includes the *H19/IGF2* cluster. The second, proposed by Smilinich *et al.* (Smilinich *et al.* 1999) includes the *KCNQ1* gene and its associated antisense transcript. A comprehensive study by Engemann *et al.* (Engemann *et al.* 2000), in which they compared the mouse and human sequences in the region of the putative second imprinting control centre, offered new insights into the cluster's overall regulation. The 1Mb imprinted region is divided into two separate domains by an interval that is predominantly non-imprinted. This region contains the genes *LTRPC5* to *TSSC6*, of which most are biallelically expressed. Of note, however, is that this region displays differences between human and mouse – imprinting is not as well conserved in this region as it is in other parts of 11p15.5, since *LTRPC5* is imprinted in human but biallelic in mouse, whereas *TSSC4* shows tissue-specific imprinting in mouse but is biallelic in human. However, Engemann's sequence comparison of this region revealed strong conservation of gene order and transcriptional orientation between the two species. A site of particularly strong conservation was identified in the 3' region of the *KCNQ1* gene, including the intronic sequences and the antisense transcript. Such an unusually high degree of similarity between mouse and human implies that this region may have a crucial role in regulation. Engemann *et al.* pursued this hypothesis in the mouse and demonstrated that the CpG island associated with the *Kcnq1* antisense transcript

(*Kcnq1ot1*) was significantly methylated in oocytes, and unmethylated in sperm. These epigenetic differences were maintained in zygotes and embryonic stem cells, indicating that this CpG island has a germline methylation imprint. The only other known germline imprint mark in this region is at the *H19* DMR, and thus it appears almost certain that the *KCNQ1OT1* CpG island has a pivotal role in the second imprinting control centre.

1.2.1.1.7 Summary

In summary, this cluster of imprinted genes is a model exemplifying some of the imprinting control mechanisms that exist in humans. The reciprocal imprinting of maternally expressed *H19* and paternally expressed *IGF2* appears to be regulated by a form of enhancer competition, with a 'switch' mechanism involving the activity of a methylation-sensitive insulator or boundary element. Biallelic expression of *IGF2*, through loss of imprinting, causes BWS, and may occur by two separate pathways, either in conjunction with *H19* hypermethylation, or completely independently of it, with accompanying loss of methylation of *KCNQ1* and biallelic expression of an antisense transcript, *KCNQ1OT1*, in the latter. Thus, by studying the breakdown of imprinting control in a disease such as BWS, at least two different mechanisms of *IGF2* imprinting regulation have been elucidated. BWS is also caused in a significant proportion of cases by mutations of the gene *CDKN1C*. The fact that BWS can arise as a result of aberrations of two distinct genes (*IGF2* and *CDKN1C*), yet cases in each of these two categories are indistinguishable, implies that these genes may act in a common pathway. Evidence to support this idea has emerged recently, first from a double knockout mouse deficient in *CDKN1C* and overexpressing *IGF2*, which indicated that the products of these genes may interact antagonistically in cell proliferation (Caspary *et al.* 1999); and second, from the experiments of Grandjean *et al.* (Grandjean *et al.* 2000), in which increased levels of IGF2 protein have been correlated both *in vitro* and *in vivo* to a decrease in *CDKN1C* expression. Future research may clarify these interactions, and reveal new insights into the causative mechanisms of BWS.

1.2.1.2 Other clusters in the human genome

There is one other known major imprinted gene cluster in the human genome, at 15q11-q13. This is discussed in 1.2.2.1, in the context of imprinting centre control, as it represents the best example of this feature. However, it is likely that there are other clusters that are as yet unidentified, and one potential candidate may be a cluster at the *PEG1/MEST* locus on chromosome 7.

1.2.1.2.1 Possible cluster at 7q32

The first identified imprinted gene on human chromosome 7 was the predominantly paternally expressed gene *PEG1/MEST* (Kobayashi *et al.* 1997; Riesewijk *et al.* 1997), which is discussed further below in 1.2.3.3, in the context of its recently demonstrated isoform-specific imprinting. To investigate whether this gene is in fact part of a new cluster of imprinted genes on this chromosome, exon-trapping experiments were employed to assess the surrounding region (Blagitko *et al.* 1999). Unknown exons were detected, and subsequently found to form part of a novel gene, *COPG2* (coatomer protein complex subunit), which encodes a protein with homology to bovine γ-COP. The gene structure was elucidated, and its relationship to the adjacent gene *PEG1/MEST* was determined. The two genes are orientated in a tail-to-tail manner, with an overlap of 52bp at the 3' utr of both genes.

Using an intronic polymorphism in intron 22, Blagitko *et al.* performed imprinting analysis of *COPG2* by RT-PCR, which revealed monoallelic expression in almost all fetal tissues, except brain and liver (Blagitko *et al.* 1999). In those tissues subject to imprinting, expression of the gene was determined to be exclusively paternal, as in *PEG1/MEST*. The finding of this imprinted gene in the vicinity of another imprinted gene thus appeared to define a new cluster of imprinted genes. Subsequent to the study of Blagitko *et al.*, however, an independent study has claimed that expression of *COPG2* is biallelic (Yamasaki *et al.* 2000). Yamasaki *et al.* demonstrated biallelic expression of the gene in fetal tissues and adult lymphocytes using a single nucleotide polymorphism located in exon 17. In addition, they offer an explanation for the discrepancy between their study, and the previous study of Blagitko *et al.*;

they identified additional 3' utr sequence for *MEST*, which extends the overlap of the two genes to include the last four exons and introns of *COPG2*, rather than just the previously described 52bp overlap. If this is the case, then, as they propose, the imprinting analysis of Blagitko *et al.* may have been complicated by the overlapping origin of the transcripts that were analysed. Indeed, Yamasaki *et al.* demonstrated in their study that the RT-PCR products of Blagitko's experiment were most likely to have been predominantly derived from *MEST*, such that their conclusions of monoallelic expression were representative of *MEST* rather than *COPG2*.

However, despite the probable biallelic expression of the *COPG2* gene, this region remains a potential imprinted domain, since Yamasaki *et al.* identified *CIT1*, an antisense transcript of *COPG2* intron 20 that was paternally expressed in fetal brain, liver and lung, but not adult lymphocytes.

1.2.2 Imprinting centres may control the switching of imprints in the germline

From studies of the cluster of imprinted genes at 11p15.5, and experimental data from investigations of imprinting mechanisms in germline development, it is apparent that the control of imprinting is to some extent due to the action of imprinting centres. The concept of a regionally acting element with the capacity to influence the imprinting of a chromosomal domain certainly gives some explanation as to why imprinted genes typically reside in clusters. Elucidating the role of such imprinting centres is currently a popular aspect of imprinting research, and perhaps the most significant progress has been made in analysing the cluster of imprinted genes at 15q11-q13.

1.2.2.1 Cluster of genes at 15q11-q13 may be under the control of an imprinting centre

Along with the cluster at 11p15.5 (see 1.2.1.1), the other most extensively studied cluster of imprinted genes is that at 15q11-q13. Aberrations or loss of imprinted

genes in this region are associated with the diseases Angelman syndrome and Prader-Willi syndrome. There is strong evidence that this cluster is regulated by the action of an imprinting centre, and this represents a prime example of the function of imprinting centres.

1.2.2.1.1 Deficiencies of imprinted gene(s) in a cluster at 15q11-q13 implicated in Angelman Syndrome and Prader-Willi Syndrome

The identification of imprinted genes has in many cases only occurred through the detection of abnormal inheritance patterns of a disease, and the subsequent localisation of the genes responsible. This was the case for the discovery of the imprinted gene cluster at 15q11-q13, which was identified through chromosomal analysis of patients with Angelman Syndrome (AS) and Prader-Willi syndrome (PWS) (Williams *et al.* 1990; Magenis *et al.* 1990). Angelman syndrome is characterised by severe developmental delay and ataxia, whereas Prader-Willi syndrome is a neuro-behavioural disorder characterised by neonatal hypotonia, developmental delay and hyperphagia (Lee, Wevrick 2000). Both diseases occur at a frequency of 1 in 15000 live births, and both are associated with large deletions of a region of chromosome 15q11-q13 in 70% of cases, but the origin of these deletions is consistently paternal in PWS and maternal in AS (Knoll *et al.* 1989). The exclusive, parent-specific inheritance associated with each disease suggests that a deficiency of at least one gene subject to genomic imprinting may be the underlying cause of each disease (Nicholls 1993).

In support of this idea, nearly all non-deletion cases of PWS are associated with maternal uniparental disomy for chromosome 15, and likewise, a smaller, yet still significant proportion (3-5%) of AS cases are associated with paternal uniparental disomy for chromosome 15 (Knoll *et al.* 1991). The parental origin of both deletions and uniparental disomy thus indicates that the absence of a paternally expressed imprinted gene(s) causes PWS, whereas the reciprocal situation of the lack of a maternally expressed imprinted gene(s) leads to AS.

1.2.2.1.2 Imprinting of genes in the 15q11-q13 cluster

Characterisation and correlation of the breakpoints flanking the deletions in PWS and AS patients has defined a common region of deletion, spanning approximately 4 megabases. Several genes reside in this region, namely a zinc-finger protein, *ZNF127*; the gene encoding the small nuclear ribonucleoprotein-associated polypeptide N, *SNRPN* (Ozcelik *et al.* 1992); *necdin* (*NDN*) (MacDonald, Wevrick 1997; Jay *et al.* 1997); and *IPW*, although this gene does not appear to have any coding capacity (Wevrick *et al.* 1994). These four genes are paternally expressed and silent on the maternally inherited allele. A further gene, *UBE3A*, which encodes a ubiquitin-protein ligase (Sutcliffe *et al.* 1997), is partially imprinted, being biallelically expressed in most tissues, but maternally expressed in brain (Vu, Hoffman 1997; Rougeulle *et al.* 1997). Three $GABA_A$ receptor genes also reside in this interval, but the imprinting of these has either not been established, or the evidence for imprinting is contradictory.

UBE3A has been proposed as a candidate gene for AS, since inactivating mutations have been found in this gene in patients with no other abnormality at this region (i.e. non-deletion/ non-UPD patients) (Matsuura *et al.* 1997; Kishino *et al.* 1997). For AS, at least, a loss of this single gene may be sufficient in itself to cause the disease. For PWS however, it is not clear whether a single locus, or the involvement of multiple genes is responsible for the phenotype. Of these two scenarios, there is stronger evidence that PWS is a contiguous gene syndrome, because no mutations have been identified that affect the expression of only one of the four paternally expressed genes in this region. It is more likely therefore that several of these genes must be disrupted for the condition to arise (Lee, Wevrick 2000). The identification of novel transcripts in this region of 15q11-q13 indicates the strong possibility of additional genes residing here (Lee, Wevrick 2000). Furthermore, a novel imprinted gene in the PWS region, named *NDNL1* (*NDN*-like 1)/ *MAGEL2* has recently been characterised, and on the basis that it is expressed predominantly in the brain, and shows paternal-specific expression, it must be investigated as a further candidate gene for PWS (Boccaccio *et al.* 1999; Lee *et al.* 2000).

1.2.2.1.3 Regulation of imprinting at 15q11-q13 is mediated through an imprinting centre (IC), of which mutations are a novel cause of AS and PWS

The imprinted cluster at 15q11-q13 is subject to a tightly maintained methylation pattern such that the maternal chromosome is predominantly methylated, and the paternal chromosome which expresses most of the genes, is unmethylated. Recently, a new subgroup of PWS and AS patients has indicated that disruptions to the establishment of this pattern may be an alternative cause of the diseases. This subgroup of patients does not carry the common 4 Mb deletions, nor do they have uniparental disomy. In fact, at the genetic level they have apparently normal, biparental inheritance of chromosome 15. However, at the epigenetic level, they are found to possess either a typical maternal methylated pattern on both chromosomes (in PWS), or a typical paternal unmethylated pattern on both chromosomes (in AS) (Buiting *et al.* 1995; Sutcliffe *et al.* 1994). In functional terms, the outcome of this is presumably comparable to having maternal or paternal uniparental disomy, respectively, and the expression pattern of the imprinted genes in this region reflects this.

Based on this finding, it was postulated that an imprinting centre might control the resetting and necessary switching of the parental imprint of each chromosome during gametogenesis. The subset of patients described above is now known to have microdeletions in the 5' region of the *SNRPN* gene, which is now proposed as the imprinting centre (IC). It is possible that mutations of the imprinting centre, in the form of microdeletions, prevent the imprint of a chromosome being reset during gametogenesis. Such mutations can therefore remain phenotypically silent for several generations, but will become phenotypically apparent when the sex of the transmitting individual changes. For example, if a female individual carries an imprinting mutation of the IC, which prevents the switching of the maternal imprint to a paternal imprint, this mutation will be passed on silently through the female germline. However, if she passes this mutation to her son, the maternal imprint cannot now be reset to a paternal one during spermatogenesis, and his resulting offspring will have a 50% chance of inheriting the chromosome that carries an eternally maternal imprint. Thus, these offspring, regardless of their sex, would

carry two copies of chromosome 15 both with a maternal imprint, and would have PWS. A similar but reciprocal situation would cause AS in the case of an individual with an imprinting mutation that prevents the switching of a paternal imprint to a maternal one. This would be passed silently through the male germline until the transmitting individual is female, at which point her offspring have a 50% chance of inheriting the mutation which will leave them with two paternally imprinted copies of chromosome 15, resulting in AS (Ohta *et al.* 1999).

Analysis of the extent of the microdeletions in several individual cases of PWS and AS IC mutations indicates that the imprinting centre at 15q11-q13 actually comprises two parts, which are separately responsible for the switching of the male or female germline imprints. The deletion mapping studies have placed the bipartite IC at the promoter region of *SNRPN*. Since it acts on genes both upstream and downstream of this locus, it has the ability to control imprinting regulation *in cis*, and in a bidirectional manner. The most recent addition to these established ideas is a suggestion that the imprinting centre has a further role of maintenance of the imprint in somatic cells (Mann, Bartolomei 2000; Bielinska *et al.* 2000). This is a novel concept, proposing that the imprinting centre is not required solely for germline establishment of the imprint, but is also crucial in the postzygotic maintenance of the epigenotype.

Full elucidation of the mechanisms of the action of this imprinting centre is yet to be completed, but the demonstration of a controlling element such as the IC in this cluster suggests new ideas about imprinting regulation that may well be applicable to the other imprinted clusters of the human genome, both those that are known, and potentially other as yet unidentified clusters.

1.2.3 Tissue-specific, promoter-specific, and isoform-specific imprinting

Most imprinted genes have been found to conform to an established pattern of imprinting, which is most often consistent with that found in the homologous mouse

gene. However, as imprinting research delves further into the finer details of imprinting regulation and mechanisms for specific genes, an increasing number of genes have been found to show some degree of tissue-specific or time-specific imprinting. This temporal and spatial variation in imprinting is accompanied by variation between mouse and human, to the extent that not all genes follow the same pattern in both species. The two most notable examples of imprinted genes that show complex isoform or promoter-specific imprinting patterns, including bidirectional imprinting within the same gene, are *GNAS1*, which has been well characterised, and *GRB10*, for which the evidence is less well established, and even contradictory. Both are discussed below, along with the genes *IGF2* and *PEG1/MEST*, which may also show promoter or isoform-specific imprinting patterns.

1.2.3.1 *GNAS1*

GNAS1, located at human chromosome 20q13.11, is a prime example of a gene showing complex promoter-specific imprinting patterns. The gene encodes the α subunit of the G protein ($G_S\alpha$), which functions to couple receptor binding by a variety of hormones, to cAMP generation (Kozasa *et al.* 1988).

1.2.3.1.1 Association with PHP-Ia

Null mutations of *GNAS1* give rise to various phenotypes, most usually causing the hormone-resistance syndrome type Ia pseudohypoparathyroidism (PHP-Ia) (Levine *et al.* 1988). In addition to hormone resistance, this condition is associated with a 50% reduction in erythrocyte G_S activity and a skeletal abnormality known as Albright hereditary osteodystrophy (Levine *et al.* 1986). A milder form of the disease, in which patients have most of the symptoms of PHP-Ia, but do not suffer hormone resistance, is known as pseudopseudohypoparathyroidism (PPHP).

The possible involvement of imprinting in this disease was suggested by the familial occurrence of PHP-Ia and PPHP within the same families. Indeed, a study of the

inheritance patterns of *GNAS1* mutations revealed that PHP-Ia was consistently the result of a maternal transmission, whereas PPHP was always associated with a paternal transmission (Davies, Hughes 1993).

1.2.3.1.2 Biallelic and paternally-derived transcripts, $G_S\alpha$, and XLαs

At first, biallelic expression of *GNAS1* was found to be prevalent in many tissues (Campbell *et al.* 1994). However, the discovery of a novel, alternative promoter and first exon of *GNAS1*, 35kb upstream of the original first exon, identified an alternative transcript that encodes the large G protein XLαs, and is expressed exclusively from the paternally derived allele (Hayward *et al.* 1998a).

1.2.3.1.3 A maternally derived transcript, *NESP55*

The discovery of both biallelic and paternally derived transcripts confirmed the idea that *GNAS1* is subject to an imprinting effect, but did not fulfill the predictions of the consistently maternal inheritance of PHP-Ia, which suggest the involvement of a maternally expressed gene. However, further searches at the *GNAS1* locus revealed the presence of an alternative promoter and first exon, 11kb upstream of the XLαs promoter (Hayward *et al.* 1998b). This component gives rise to a transcript that, like XLαs, splices onto *GNAS1* exon 2, yet the resulting proteins for each promoter are structurally unrelated. The novel 5' exon gives rise to a transcript that is maternally expressed, and encodes a neuroendocrine secretory granule protein NESP55.

1.2.3.1.4 Complex splicing and imprinting patterns of *GNAS1*

The imprinting of *GNAS1* is not simple. As described above, the gene comprises biallelic, paternally and maternally expressed variants, in which the nature of imprinting is determined by the use of independent promoters and alternative splicing arrangements. There are several other complexities to the *GNAS1* situation that are not discussed here, including further splicing variations and exon skipping events that yield a multitude of different isoforms, and the existence of a fourth

potentially imprinted alternative promoter (exon 1A) (Liu *et al.* 2000a). Additionally, an antisense transcript (discussed in 1.2.4) may be involved in imprinting regulation (Hayward, Bonthron 2000; Wroe *et al.* 2000). See Figure 3.

1.2.3.2 *PEG1/MEST*

The mouse *Peg1* gene had been isolated in a subtractive hybridisation screen for novel paternally expressed genes (Kaneko-Ishino *et al.* 1995), and was found to be identical to the previously identified gene *Mest*, the protein product of which shows homology to the alph/beta-hydrolase fold family, and is expressed most abundantly in the early embryo (Nishita *et al.* 1996). The human homologue, *PEG1/MEST*, has been located at the equivalent human chromosomal region of 7q31.3, and found to be imprinted, as discussed below.

1.2.3.2.1 Imprinting of *PEG1/MEST*

On the basis that imprinting is generally conserved between mouse and human, establishing the imprinting status of the human *PEG1/MEST* was a natural area for research, particularly as no imprinted genes were as yet known to reside on this chromosome. Two groups demonstrated predominantly monoallelic, paternal expression of *PEG1/MEST* (Kobayashi *et al.* 1997; Riesewijk *et al.* 1997), and another investigated the methylation status of the *PEG1* promoter and found parent-specific methylation, with the paternal allele consistently unmethylated, and the maternal allele consistently methylated (Lefebvre *et al.* 1997).

Although the majority of the imprinting experiments indicated that *PEG1/MEST* follows a strict pattern of paternal monoallelic expression, accompanied by maternal methylation of the promoter, one exception to this consensus was detected by Riesewijk *et al.* (Riesewijk *et al.* 1997), who found biallelic expression of the gene in adult lymphocytes. This observation was substantiated by the same group, through the detection of *PEG1/MEST* transcripts in a lymphoblastoid cell line with maternal uniparental disomy for chromosome 7. The lack of any paternal chromosome 7

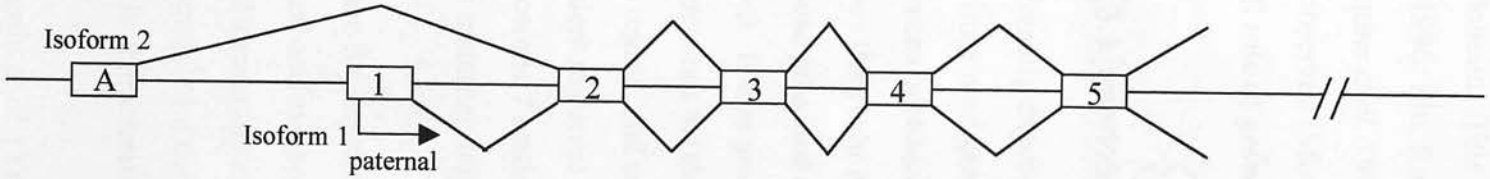
contribution in this cell line demands that any expression of the gene must therefore be maternally derived.

1.2.3.2.2 Isoform-specific imprinting of *PEG1/MEST*

The discovery of apparently biallelic expression of *PEG1/MEST* in lymphocytes (Riesewijk *et al.* 1997) has prompted further investigation of the mechanisms that cause this loss of imprinting in some tissues. Kosaki *et al.* (Kosaki *et al.* 2000) assembled ESTs for the 5' region of *PEG/MEST*, and identified a subgroup that included a novel first exon. Indeed, analysis of the genomic region confirmed that an alternative first exon, named exon A, was located 6kb upstream of the original exon 1. Alternative splicing of these two exons onto the remaining exons creates two different splice variants of the gene. See Figure 3. The most significant finding is that the newly identified isoform of this gene is subject to biallelic rather than monoallelic expression. Interestingly therefore, it is not the supposed loss of imprinting of the gene in lymphocytes that causes biallelic expression. Rather, it is the fact that lymphocytes express both isoforms of this gene, one of which is paternally expressed, and the other biallelically expressed. The previous experiments, not taking account of this fact, drew their conclusions from RT-PCR experiments using primers that were not able to discriminate between isoforms. The paternal, monoallelic expression of the original isoform was therefore masked by the biallelic expression of the novel isoform in Riesewijk's study.

What was previously assumed to represent tissue-specific imprinting of *PEG1/MEST* is now explained by isoform-specific imprinting. As pointed out by Kosaki *et al.* (Kosaki *et al.* 2000), the finding of biallelic expression for other genes may have overlooked the possibility that opposite or biallelic imprinting of different isoforms of the same gene can mask the results. A re-evaluation of imprinting in genes previously thought to be biallelically expressed, particularly in cases where the mouse homologue is subject to imprinting, may yield some interesting findings.

a) *PEG1/MEST*



b) *GNAS1*

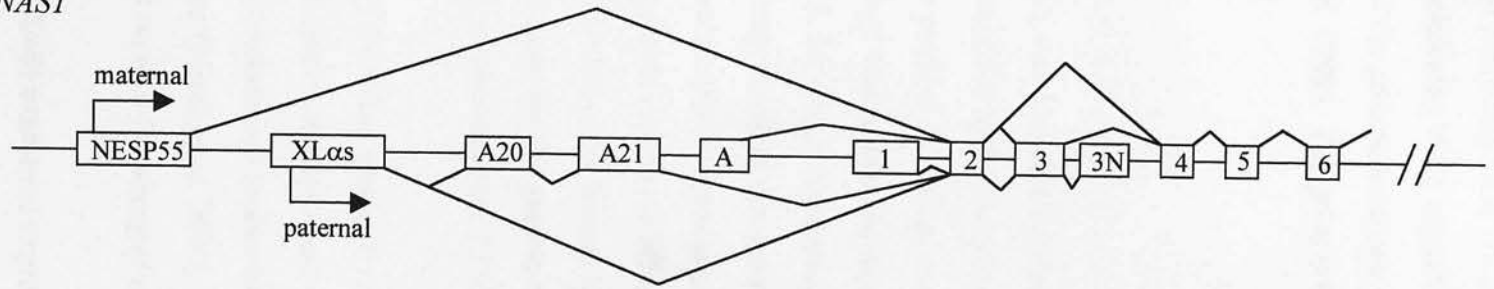


Figure 3. Examples of isoform and promoter-specific imprinting. a) *PEG1/MEST* has two known isoforms, of which the first (isoform 1) is paternally transcribed, but the second (isoform 2) is biallelically transcribed from an alternative upstream first exon, A. Both exons 1 and A splice onto exon 2. b) *GNAS1* has a very complex splicing arrangement. Transcription begins from one of three alternative first exons, NESP55, XLαs, and exon 1, and is maternal, paternal or biallelic, respectively. Further complexity is present in the use of alternative exons, including A20 and A21. Exon skipping events result in additional isoforms. A further exon, A, is involved, and may be imprinted. See text for discussion. Diagram is not to scale. In both cases, downstream exons are not shown. Redrawn after Kosaki *et al.* (2000), and Hayward *et al.* (2000).

1.2.3.3 *GRB10*

The gene *GRB10* encodes a growth factor receptor binding protein, which has the capacity to bind both insulin and insulin-like growth factor receptors by means of a SH2 domain. This has the effect of inhibiting their tyrosine kinase activity (O'Neill *et al.* 1996; Liu, Roth 1995), and inhibiting IGF1 signalling of cell proliferation (Morrione *et al.* 1997). Binding of the growth hormone receptor (GHR) has also been observed (Moutoussamy *et al.* 1998). The gene is therefore believed to have an overall role of growth suppression.

1.2.3.3.1 Imprinting of *GRB10*

A maternally expressed gene, *Meg1*, was identified during a systematic screen for novel imprinted genes, using a subtractive hybridisation approach to detect differences between the expression profiles of normal and androgenetic mouse embryos (Miyoshi *et al.* 1998). *Meg1* was in fact identical to the known gene *Grb10* on mouse proximal chromosome 11, for which imprinting had not previously been reported. It was proposed that the imprinting of this gene, in addition to its function, could account for the growth promotion effects associated with paternal duplications of this region, and conversely, the growth repressive effects associated with equivalent maternal duplications. Similarly, the human homologue of *Grb10* on chromosome 7 could be a candidate gene for the growth retardation observed in human maternal uniparental disomy of chromosome 7 (Miyoshi *et al.* 1998).

Evidence for imprinting of human *GRB10* has emerged recently, but is somewhat complex, and even contradictory. Initially, tentative and speculative evidence for *GRB10* imprinting came from the asynchronous replication timing that was observed, as compared to a non-imprinted gene (Monk *et al.* 2000). Asynchronous replication timing is characteristic of imprinted regions (Kitsberg *et al.* 1993).

Yoshihashi *et al.* (Yoshihashi *et al.* 2000) examined expression of *GRB10* in fetal tissues and found exclusively monoallelic expression in fetal brain. To determine the

parental origin of the expressed allele they analysed somatic cell hybrids in which the human component was derived from human lymphocytes, and concluded that the *GRB10* gene was only expressed from the maternally derived chromosome. However, the imprinting of *GRB10* may not be as simple or straightforward as this. The experiments of Yoshihashi *et al.* (Yoshihashi *et al.* 2000) assume that imprinting of the gene is consistent between different tissues, that it is faithfully maintained in the somatic cell hybrids, and do not take into account the potentially splice-variant specific imprinting such as that which has been observed in *GNAS1*, and discussed above. Indeed, the most comprehensive analysis of *GRB10* has taken these factors into consideration, and has found them to be of significance with regard to the imprinted status of *GRB10* (Blagitko *et al.* 2000). Blagitko *et al.* found biallelic expression in all fetal tissues, except brain, in which expression was solely paternal. When they extended their analysis to examine the expression status of each of the four splice variants separately, one isoform, *GRB10γ1* displayed monoallelic expression exclusively from the maternal allele in the two fetuses analysed. These experiments thus present evidence for both biallelic, and monoallelic expression of *GRB10*, with bidirectional imprinting of the latter case, depending on the tissue, and the splice variant involved (Blagitko *et al.* 2000).

1.2.3.3.2 Silver-Russell Syndrome and the possible involvement of an imprinted gene on chromosome 7

Silver-Russell syndrome (SRS) is a disease characterised by prenatal and postnatal growth retardation, and dysmorphic features including triangular facies and fifth finger clinodactyly (Silver *et al.* 1953; Russell 1954). A genetic basis to SRS is indicated by the presence of maternal uniparental disomy in approximately 7% of cases (Preece *et al.* 1997; Kotzot *et al.* 1995). The fact that paternal uniparental disomy of chromosome 7 does not cause SRS (Hoglund *et al.* 1994) specifically indicates non-equivalence of the parental chromosomes, and is suggestive of an imprinting effect. As with other diseases associated with a uniparental disomy, it is inconclusive from the maternal UPD cases alone whether the disease is caused by the resultant over-expression of a maternally expressed growth-suppressing imprinted

gene, or the effective loss of a paternally expressed growth-promoting imprinted gene on this chromosome. The phenotypic effects of both scenarios could result in the growth retardation observed in SRS patients, and may be indistinguishable from each other.

It seems probable, therefore, from the significant proportion of SRS cases with maternal UPD of chromosome 7, and the other evidence mentioned above, that an imprinted gene on chromosome 7 has a fundamental involvement in this disease. Of note is that the possibility of SRS being caused by the effects of a recessive gene (with two recessive copies exposed by maternal UPD7) is eliminated due to the finding of uniparental heterodisomy cases, in which the two copies of chromosome 7 are derived from the same parent, but are not identical, unlike uniparental isodisomy.

Whilst the evidence described earlier regarding the imprinted status of *GRB10* is complex, and variable between studies, it is apparent that there is at least some maternal specific expression of this gene, which could be envisaged to cause an abnormal and increased amount of the protein product as a result of maternal UPD7. Since the protein has a growth suppressing function, the resulting effects could feasibly account for the growth retardation observed in SRS. As an imprinted gene therefore, *GRB10* is a good candidate. Its candidacy is further strengthened by the discovery of small duplications of 7p11-p13 by Joyce *et al.* (Joyce *et al.* 1999) and Monk *et al.* (Monk *et al.* 2000), which, in conjunction with the maternal UPD7 cases, clarify for the first time that it is the over-expression of a maternally expressed gene, rather than the absence of a paternally-expressed gene, that causes SRS. Joyce *et al.* reported an inverted, interstitial duplication of 7p12.1-p13, including the genes *GRB10* and *IGFBP1* (insulin-like growth factor binding protein 1) in a mother and daughter, both of whom had characteristic features of SRS. Monk *et al.* identified a similar duplication of 7p11.2-p13, also including *GRB10*, *IGFBP1* and *IGFBP3*, in another child with SRS. The genes *IGFBP1* and *IGFBP3* might be considered candidates for the SRS gene on the basis of their chromosomal location in these small duplications, but are unlikely to be involved as they are known to be biallelically expressed (Wakeling *et al.* 2000).

With regard to the vast majority of Silver Russell cases, which are apparently karyotypically normal, and do not possess uniparental disomy or duplications, alternative explanations must be identified for the presence of the disease in these patients. If upregulation of *GRB10*, by either maternal uniparental disomy, or maternally derived duplications, is the underlying cause of SRS, then there are three remaining theoretical routes by which it could also be functionally upregulated. First, there may be submicroscopic duplications present, which are undetectable by cytogenetic methods. Second, the gene may acquire gain of function mutations, which cause an upregulation of its expression, and third, it may suffer loss of imprinting or imprinting mutations, such that both copies take on a maternal epigenotype, with a similar outcome to maternal UPD or duplications.

A search for mutations of *GRB10* was conducted by Yoshihashi *et al.* (Yoshihashi *et al.* 2000) in 30 Japanese and 28 European SRS patients. Two of the patients were found to have a point mutation resulting in a proline to serine substitution in the amino-terminal region of the GRB10 protein. In both cases the mutation was also present in the mother. Grandparental material was available for one case, showing that the mother of the proband had inherited the mutation paternally, and thus offering an explanation as to why she did not have symptoms of SRS. The significance of the parental origin of the mutation, and its absence from normal matched controls suggests that it may be important in the disease. However, an understanding of the functional effect of the mutation on the protein, and the mechanism by which it could upregulate expression of *GRB10* to cause increased growth suppressing activity is yet to be elucidated.

Whilst *GRB10* is a compelling candidate gene for SRS, another imprinted gene on chromosome 7, *PEG1/MEST*, at 7q32, has not been excluded. Indeed, the possibility that the long arm of chromosome 7 carries the SRS gene cannot be dismissed because of an interesting case of a patient with SRS who carried paternal isodisomy of 7p, and maternal isodisomy of 7q (Eggerding *et al.* 1994); until the identification of this patient all previous uniparental disomies had encompassed the whole of

chromosome 7. Whilst this case in itself cannot determine whether it is the loss of a paternally expressed gene or the excess of a maternally expressed gene that is the cause of SRS, it does imply that the gene is located on the long arm of chromosome 7. As the first identified paternally expressed imprinted gene on chromosome 7, *PEG1/MEST* (discussed earlier in this chapter in the context of clustering and isoform-specific imprinting), was proposed as a candidate for SRS. Given that *PEG1/MEST* is paternally expressed, it would be expected that a loss of function of this gene would be related to SRS. Thus, in cases of SRS that do not carry maternal uniparental disomy, inactivating mutations might represent an alternative causative mechanism. However, in a mutation screen of 49 SRS patients, only three silent nucleotide alterations were detected, and methylation patterns of the 5' region of *PEG1/MEST* were found to be normal in 35 SRS patients (Riesewijk *et al.* 1998). The evidence for the involvement of *PEG1/MEST* in SRS is weak, but a recent and very significant study has renewed the interest in 7q; Hannula *et al.* (Hannula *et al.* 2001) detected segmental maternal UPD of chromosome 7, extending from 7q31-qter only. The remainder of the chromosome displayed biparental inheritance. Other than the patient mentioned above with maternal uniparental disomy for the entire long arm of chromosome 7, this is the first report of segmental maternal UPD7. The 35Mb segment thus represents a candidate SRS interval, and contains *PEG1/MEST* and *COPG2*, but excludes *GRB10*. Focussing on this potential cluster may identify further genes, which if imprinted may be considered SRS candidates.

The evidence regarding the location of the SRS gene is confusing and in some ways equivocal in that there has been convincing evidence for both the 7p11-p13 locus (containing *GRB10*), and the 7q32 locus (containing the putative cluster that includes *PEG1/MEST*). It is not impossible that there is more than one locus or gene responsible for SRS; indeed, given that the disease itself is so clinically heterogeneous, perhaps this reflects genetic heterogeneity. There may be two or more distinct genes that result in similar phenotypes. Alternatively, there may be an as yet unelucidated reason to explain why the evidence has been so contradictory.

1.2.3.4 Promoter specific imprinting of *IGF2*

In general, *IGF2* is considered to be a paternally expressed imprinted gene. However, like *GNAS1*, *GRB10* and *PEG1/MEST*, *IGF2* has a slightly more complex expression pattern, which is related to the use of alternative promoters. *IGF2* is expressed monoallelically, and exclusively from the paternal allele, in fetal tissues, including liver, but biallelic expression is observed in adult liver (Kalscheuer *et al.* 1993). However, rather than this apparently biallelic expression occurring as a result of loss of imprinting, it is proposed that the gene *IGF2* utilises one of four different promoters P1-P4, and when the promoter P1 is used (as in adult liver), expression of the gene is biallelic. Utilisation of all other promoters results in expression of the paternal allele only (Vu, Hoffman 1994). The conclusion from this example is that a gene can be imprinted or non-imprinted in a single tissue, depending on which promoter is used.

1.2.4 Antisense transcripts may be involved in imprinting regulation

An increasing number of antisense RNA transcripts associated with imprinted genes are being discovered, challenging traditional ideas of the mechanisms by which imprinted genes are regulated. These antisense transcripts are also imprinted, sometimes coordinately and sometimes reciprocally with respect to the imprinted status of the genes with which they are associated. Recent studies of the currently small, but increasing number of antisense RNA transcripts suggest that they may indeed possess a genuine and novel role in imprinting control.

1.2.4.1 Antisense transcripts identified to date

1.2.4.1.1 Mouse *Igf2r*

The imprinted mouse *Igf2r* gene contains two differentially methylated regions. The first of these, located upstream of the transcriptional start site, is correlated with

expression of the *Igf2r* gene, in that the paternal allele is methylated at this region, and is transcriptionally suppressed, whereas the transcriptionally active maternal allele is unmethylated. However, the second differentially methylated region, located in the second intron of *Igf2r*, is methylated in a reciprocal manner, such that the maternal allele is methylated and the paternal allele is unmethylated. This second differentially methylated region is now known to be the site of a promoter for an RNA transcript that is transcribed in a paternal-specific manner from the antisense strand (Wutz *et al.* 1997).

Analysis of the imprinting of the sense and antisense transcripts at the *Igf2r* locus has drawn parallels with the reciprocally imprinted gene pair *H19* and *Igf2*, in the sense that an expression competition model may be involved. The experiments of Wutz *et al.* (Wutz *et al.* 1997) showed that deletion of the second region causes loss of imprinting of the sense transcript altogether, resulting in biallelic expression. Their explanation is that the reciprocally imprinted overlapping antisense RNA is transcribed from the paternal chromosome, (which is unmethylated at DMR2), and this transcription prevents expression of the paternal sense transcript. Deletion, therefore, eliminates this element of control from the system, allowing transcription of both parental alleles from the sense gene.

The actual mechanisms by which the antisense exerts its negative regulatory effect on the sense gene are yet to be deduced, but possibilities are that it somehow occludes the *Igf2r* promoter to prevent binding by transcription complexes, that it causes local inactivation of the chromosome by binding to the chromatin, or that there is competition between the sense and antisense promoters for shared enhancers or transcription factors (Reik, Constancia 1997).

1.2.4.1.2 *GNAS1*

The extraordinary imprinting pattern of the *GNAS1* gene, encoding maternal (*NESP55*), paternal (*XLαs*) and biallelic ($G_S\alpha$) transcripts has been detailed previously (1.2.3.1). However, a further feature of this complex gene is the existence

of a spliced, paternally expressed antisense transcript, which originates in the differentially methylated region of *XLαs*, and is transcribed through the upstream, oppositely imprinted sense *NESP55* exon (Wroe *et al.* 2000; Hayward, Bonthron 2000). Antisense transcription through sense exons of the associated gene is a frequent feature of antisense RNAs, and given the close proximity in this case, it is conceivable that transcription of the antisense extending through the *NESP* exon and promoter could be a unique *in cis* suppression mechanism, to silence the paternal copy of *NESP* (Hayward, Bonthron 2000).

1.2.4.1.3 *KCNQ1*

The maternally expressed gene *KCNQ1* in the imprinted cluster at 11p15.5 is associated with a reciprocally imprinted, overlapping antisense transcript, *KCNQ1OT1*, which is paternally expressed (Mitsuya *et al.* 1999; Smilinich *et al.* 1999). As with mouse *Igf2*, an intronic CpG island at *KCNQ1* is subject to maternal-specific methylation, which apparently dictates the paternal-specific expression of the antisense transcript. Unlike the *GNAS1* antisense, this antisense transcript is not spliced, and is thought to span approximately 60kb.

Of significant note is that loss of maternal methylation at this DMR is linked to Beckwith-Wiedemann Syndrome, and may represent an alternative mechanism by which the disease can arise. Studies have demonstrated that loss of imprinting and resultant biallelic expression of the antisense transcript occurs in a large proportion or even the majority of BWS patient samples analysed (Mitsuya *et al.* 1999; Smilinich *et al.* 1999; Lee *et al.* 1999b). How does biallelic expression of the *KCNQ1OT1* antisense transcript lead to BWS? One possibility is that the loss of maternal imprinting that leads to the expression of the normally silent maternal copy of the antisense results in silencing of a nearby maternally expressed gene, through mechanisms similar to the competition model. A candidate gene for this effect would be *CDKN1C*, as it is known that mutations of this gene alone are another cause of BWS. Maternal expression of the *KCNQ1OT1* antisense could create, and win, the competition with the upstream gene *CDKN1C*, such that the latter gene is

silenced, resulting in BWS. This idea is supported by the elegant experiments of Horike *et al.* (Horike *et al.* 2000), in which they generated targeted deletions of the *KCNQ1OT1* CpG island. The effects of this were the elimination of *KCNQ1OT1* expression from the paternal chromosome, and the activation of the normally silent paternal alleles of the genes *KCNQ1* and *CDKN1C*. The significance of this is that the loss of imprinting of *KCNQ1OT1* would result in biallelic expression, and based on the findings of Horike *et al.*, this could be expected to inactivate *CDKN1C*, and may therefore be a novel underlying mechanism of BWS.

1.2.4.1.4 *UBE3A*

UBE3A is a gene within the cluster at 15q11-q13, and is expressed maternally in the brain, but biallelically in some other tissues. An antisense transcript spanning at least the 3' half of the gene has been identified, and found to be oppositely imprinted, displaying paternal specific expression in brain (Rougeulle *et al.* 1998). The authors of this study speculate that the orientation, location and expression of the sense and antisense elements lend themselves to a competition model, as described earlier. Indeed, the fact that tissues lacking the antisense transcript show biallelic expression of *UBE3A*, is in accordance with this idea.

1.2.4.1.5 *ZNF127*

Within the Prader-Willi Syndrome region of human chromosome 15 there is a gene *ZNF127*, for which paternal, monoallelic expression has been found. Both this gene and its mouse homologue *Zpf127* are associated with an overlapping antisense transcript, *ZNF127AS/Zpf127as* (Jong *et al.* 1999a; Jong *et al.* 1999b). However, unlike the examples described above, the evidence in this case indicates that the antisense transcript is coordinately rather than reciprocally imprinted with respect to the sense gene. Both genes are therefore paternally expressed, and interestingly the expression patterns differ, such that expression of sense and antisense transcripts is mutually exclusive, but in a tissue-specific, rather than allele-specific fashion. One postulated idea for the mechanism in these circumstances is that transcription from

the sense strand itself prohibits transcription from the antisense strand (Cattanach *et al.* 1997).

1.2.4.1.6 *IGF2*

Similarly, for the mouse gene *Igf2*, several associated imprinted transcripts have been identified. Three of these, which are located in a possible region of imprinting control upstream of *Igf2*, are transcribed in the antisense orientation (Moore *et al.* 1997), and like the sense gene are maternally repressed. The function of these transcripts and the significance of their antisense orientation and co-ordinate imprinting is as yet unconfirmed, but it seems likely that they possess some mechanistic role in the regulation of imprinting of *Igf2*, and perhaps other genes in the same domain.

1.2.4.2 Possible mechanisms of antisense regulation

Gaining an understanding of the functions of antisense RNAs and their role in imprinting regulation is the subject of much current and future research. The generalistic and preconceived idea that differential methylation is the sole underlying mechanism of control must now be challenged, to integrate new concepts which may be of equal importance. A role for oppositely orientated, imprinted antisense transcripts seems increasingly likely, but the actual mechanisms remain to be elucidated. A competition model, similar to that proposed for the *H19/IGF2* genes on chromosome 11p15.5, is feasible, and could account for antisense transcripts that are reciprocally imprinted in relation to their associated sense gene. This category includes the majority of the known antisense/sense pairs, but two exceptions are *Igf2* and *Zpf127*, for which the antisense transcript is apparently imprinted in the same direction as the sense gene. An alternative mechanism must be operating in these cases.

1.3 IMPRINTING AND CANCER

A large proportion of known imprinted genes for which a function has been established are involved in embryonic growth and development or cell cycle control. This common feature is unlikely to be merely coincidental, and thus brings to light ideas that may relate to the evolution of imprinting. Accompanying this however is a cost – the role of imprinted genes in developmental pathways is crucial, yet the outcome of any aberration in these pathways could lead to a loss of developmental or cell cycle control, paving the way for tumorigenesis.

It is well established that cancer may be a consequence of sequential genetic changes such as mutations and deletions of growth-regulating genes. The discovery of an increasing number of cancer-related imprinted genes now suggests that epigenetic changes may be of more significance than previously believed.

1.3.1 Evidence for a role of imprinting in cancer

The non-equivalence of the two parental genomes due to the presence of genomic imprinting explains the inability of parthenogenetic and androgenetic embryos to survive. The role of imprinting in development is clearly paramount. Furthermore, the possibility that genomic imprinting may play a role in cancer is indicated by the existence of two types of tumour: hydatidiform mole, and ovarian teratoma. Complete hydatidiform mole is an abnormal conceptus usually resulting from the fertilization of an enucleate egg by a single haploid sperm which duplicates (Lawler *et al.* 1979), or by two haploid sperm (Ohama *et al.* 1981). In either case, its genetic constitution is diploid, and purely androgenetic (Kajii, Ohama 1977). Hydatidiform mole is characterised by a relatively well-developed trophoblast, but very little embryonic tissue. In terms of genetic background, the equivalent but opposite scenario is benign ovarian teratoma, which arises from a parthenogenetically activated egg, and therefore has only a maternal genomic contribution (Linder *et al.* 1975; Ohama *et al.* 1985). This type of tumour is characterised by a poorly developed trophoblast, but contains relatively well-developed embryonic tissue.

Both types of tumour represent the extreme consequences of complete disomy for the entire genome, (which is of paternal origin in the case of hydatidiform mole, and maternal origin in ovarian teratoma), and thus exemplify the force of genomic imprinting in preventing uniparental conception. The message from this situation is that aberrations to normal genomic imprinting can result in major developmental errors, of which tumorigenesis is a possible consequence.

Further evidence that imprinting may play a role in cancer is observed in the preferential loss or gain of chromosomal regions of specific parental origin in some cancers. In particular, Wilms' tumour shows preferential loss of the maternal chromosome 11 (Schroeder *et al.* 1987; Williams *et al.* 1989; Mannens *et al.* 1988), rhabdomyosarcoma shows preferential paternal inheritance of the disomic regions of chromosome 11p (Scrable *et al.* 1989), and neuroblastoma (see Chapter 3) is characterised by preferential maternal loss of 1p (Caron *et al.* 1993). These observations imply a role for imprinting in these diseases, because of the non-equivalence of the parental chromosomal regions involved in each case.

1.3.2 Different ways that imprinted genes may be involved in cancer

Imprinted genes are particularly vulnerable components of developmental and cell cycle pathways, since they are, by definition, functionally haploid. Therefore, the loss of an imprinted tumour suppressor gene or the upregulation of an imprinted growth-promoting gene would require only a single event. These are the two main ways in which imprinted genes are implicated in tumorigenesis. In addition, defects in imprinting centres can be envisaged to cause disruption to the imprinted status of a gene, with possible tumorigenic consequences. These ideas are discussed below in the context of some well-characterised examples.

1.3.2.1 Gain of growth promoting genes

The overexpression of a gene that has a role in promoting cell growth and division will undoubtedly lead to a disruption of control, and inappropriately increased cell growth. Tumorigenesis is a likely consequence under these conditions. Theoretically, if a growth-promoting gene is imprinted and expressed from only one parental allele in the normal situation, it is vulnerable to events that can result in its overexpression. These events include firstly a loss of imprinting which would cause biallelic expression of the gene, and secondly, uniparental disomy for the chromosome carrying the expressed allele which would result in a similar outcome due to the doubled gene dosage. At least two imprinted growth-promoting genes, *IGF2* and *PEG1/MEST* are now thought to play a role in cancer when upregulated by these mechanisms, and are discussed below.

1.3.2.1.1 *IGF2* and cancer

The cluster of genes at 11p15.5 has been implicated in cancer since the discovery of loss of heterozygosity (LOH) of this region in Wilms' tumours (Reeve *et al.* 1989; Mannens *et al.* 1988). Evidence for an imprinting effect was suggested by the non-random, parental-specific origin of these losses (Schroeder *et al.* 1987), which were preferentially maternal. Additionally, Beckwith Wiedemann syndrome (BWS), and its associated predisposition to embryonal tumours are known to be linked to uniparental disomy of chromosome 11, suggestive of an imprinting effect.

Because of the LOH of this region in Wilms' tumour, it had been expected that a tumour suppressor gene might reside in this region. Early studies were focussed around this concept, predicting that the loss of a gene with tumour suppressing properties might be involved. However, unexpectedly and in contrast to this, biallelic expression of the *IGF2* gene was detected more frequently than LOH of the region. Rainier *et al.* (Rainier *et al.* 1993) found that 69% of Wilms tumours that did not have LOH displayed biallelic expression of *H19* or *IGF2*, thus highlighting loss of imprinting (LOI) as a novel, epigenetic change associated with cancer. At the same time, Ogawa *et al.* (Ogawa *et al.* 1993) found similar evidence for LOI and

biallelic expression of *IGF2* in Wilms tumours as a new mechanism underlying the development of cancer.

From these early studies, the focus turned to investigating the extent of LOI of *IGF2* and *H19* in cancers, both fetal and adult. Biallelic expression of *IGF2* and/or *H19* has been reported in Wilms' tumours (Sullivan *et al.* 1999); renal cell carcinoma (Oda *et al.* 1998; Nonomura *et al.* 1997); hepatoblastoma (Rainier *et al.* 1995; Kim, Lee 1997; Li *et al.* 1995); Ewing's sarcoma (Zhan *et al.* 1995); head and neck squamous carcinoma (el-Naggar *et al.* 1999); ovarian cancer (Chen *et al.* 2000; Kim *et al.* 1998); meningiomas (Muller *et al.* 2000); lung cancer (Kondo *et al.* 1995; Suzuki *et al.* 1994); testicular germ cell tumours (van Gurp *et al.* 1994); breast cancer (McCann *et al.* 1996); cervical cancer (Douc-Rasy *et al.* 1996; Wu *et al.* 1997a); gastric adenocarcinoma (Wu *et al.* 1997b); colorectal cancer (Kinouchi *et al.* 1996); and acute myeloid leukaemia (Wu *et al.* 1997c).

Each of these studies has provided increasingly strong evidence that LOI of these genes may play a significant role in the development of cancer. However, the results are also confusing and at times conflicting. Some studies report LOI of *IGF2* only, whilst others report LOI of both *IGF2* and *H19* within the same tumour, or LOI of *H19* only. Establishing a true picture of the involvement of these imprinted genes is therefore no easy task. One emerging pattern however, is that the results of the studies fall into two main groups; firstly, LOI and biallelic expression of *IGF2* has been frequently reported to occur in conjunction with hypermethylation or loss of expression of *H19*, suggesting a mechanistic link between imprinting of the two genes such that biallelic expression of *IGF2* is related to downregulation of *H19* expression. Indeed, this idea is consistent with both the enhancer competition model for regulation of *H19* and *IGF2*, and also with the idea that *H19* plays a role as a tumour suppressor gene. LOI of *IGF2* in conjunction with LOI of *H19* has been observed particularly in Wilms' tumours (Steenman *et al.* 1994).

The second emerging pattern is an observed LOI of *IGF2* that is <u>not</u> associated with any downregulation of the *H19* gene, but rather, occurs independently of it. In these

cases, *H19* maintains normal monoallelic expression whilst *IGF2* suffers LOI and biallelic expression. This situation is clearly not consistent with the enhancer competition model. The cancers falling into this category are the remaining non-Wilms' embryonal tumours Ewing's sarcoma, hepatoblastoma and rhabdomyosarcoma, and the majority of the adult tumours mentioned above. This uncoupling of the imprinting of *H19* and *IGF2* is reminiscent of one of the underlying genetic mechanisms of BWS. As described previously, LOI of *IGF2* in BWS occurs either with *H19* hypermethylation, or independently of it. The latter of these two situations is also associated with maternal demethylation of the *KCNQ1* gene and is likely to be related to the biallelic expression and negative regulatory role of the antisense gene *KCNQ1OT1*. Thus, from studies of BWS it is proposed that there are likely to be at least two separate imprinting centres involved in regulation of *IGF2* imprinting, of which one is associated with the downstream gene *H19*, and the other is related to the imprinting of the upstream genes *KCNQ1*, *KCNQ1OT1* and *CDKN1C*. The results of *IGF2* and *H19* analyses here suggest that a similar scenario is true of the role of imprinting in cancer. In some cancers, particularly Wilms' tumour, LOI of *IGF2* is strongly mechanistically linked to a downregulation of *H19* expression. In a second subset of cancer types however, the LOI of *IGF2* is not associated with any changes in *H19* imprinting or expression, and thus suggests an alternative *H19*-independent mechanism of *IGF2* imprinting control, and argues against a tumour suppressor role for *H19*. In the case of BWS, the *H19*-independent LOI has now been linked to altered regulation of other genes in the cluster such as *KCNQ1* and *KCNQ1OT1*.

Nearly all studies regarding LOI in cancer have focussed on only *H19* and/or *IGF2*. However, a valuable line of research to follow now would be to assess the effect of LOI of *IGF2* in cancer on other genes in the cluster, particularly in cases where *H19* is unaffected. It may be that in these cases, the LOI of *IGF2* could now be associated with epigenetic changes in other genes, in accordance with the postulated ideas regarding a second imprinting centre. Indeed, a recent study of imprinting in hepatocarcinoma (Schwienbacher *et al.* 2000) has correlated aberrant imprinting of *IGF2* and *CDKN1C* with loss of maternal specific methylation at the differentially

methylated region of *KCNQ1*. The observed loss of imprinting at this DMR, which is normally maternally methylated and associated with a paternally expressed antisense transcript, *KCNQ1OT1*, was linked to silencing of the genes *CDKN1C* and *IGF2*. Thus, a loss of imprinting at the *KCNQ1* DMR may have a more widespread negative effect on the expression of other genes in this imprinted domain, and may be important in tumorigenesis.

One further point of interest is that LOI of *IGF2* in one study was accounted for by transcription from the biallelic promoter P1 (Chen *et al.* 2000a). Transcription from this promoter is biallelic and occurs normally only in liver. The finding of a switch in promoter usage in epithelial ovarian cancer in this study may in fact be a more widespread phenomenon, and warrants further study to determine whether this is an alternative mechanism of upregulation of *IGF2*.

1.3.2.1.2 *PEG1/MEST*

A high frequency of LOH of the chromosomal region 7q31-32 has been reported in breast cancer, suggesting that the loss of a tumour suppressor gene located in this region might be important. However, the imprinted gene *PEG1/MEST* located in this interval has recently been shown to undergo frequent loss of imprinting in invasive breast cancer, resulting in biallelic expression (Pedersen *et al.* 1999). The precise function of *PEG1/MEST* is not known, although the protein shares amino acid homology with the alpha/beta hydrolase fold family. Pedersen *et al.* suggest therefore that the upregulation of the gene function that would be expected to occur when the normally silent maternal allele is activated could result in degradation of the extracellular matrix during tumour development (Pedersen *et al.* 1999).

1.3.2.2 Loss of imprinted tumour suppressor genes

In simplified terms, normal cell growth is controlled by a finely balanced interaction between suppression of cell division, and promotion of cell growth and proliferation. Tumour suppressor genes are usually involved in inhibiting the cell cycle, or

inducing apoptosis, and are therefore antiproliferative. When the function of a tumour suppressor gene is lost however, the balance is disturbed, and can lead to tumour formation. In the 1970s, Knudson and others proposed an elegant two 'hit' model to account for the development of retinoblastoma in both sporadic and familial cases (Knudson 1971). In the familial form of the disease, one mutant copy of the retinoblastoma gene (*RB1*) is inherited. On acquiring one further mutation in a single cell, the function of *RB1* is effectively eliminated, leading to tumour formation. Thus, the inherited mutation forms the first 'hit', and a subsequent, spontaneously occurring mutation forms the second 'hit'. In sporadic cases of retinoblastoma the model still applies, with the main difference being that both 'hits' are independently acquired somatic mutations.

The Knudson model carries great significance in the study of imprinting and its relevance to cancer. Imprinted genes by definition already carry one Knudson 'hit' in the form of the imprinted and usually methylated allele that is functionally silent. This is reminiscent of the inherited mutation in the retinoblastoma scenario. Thus, only one spontaneous hit is required to inactivate the second allele, so in the case of imprinted tumour suppressor genes, there is a greater vulnerability and a much higher chance of a complete loss of expression through the two hit model than for a non-imprinted tumour suppressor gene. The second hit to inactivate the functional allele of an imprinted suppressor may be a mutation or a deletion, but loss could also occur through hypermethylation of a normally differentially methylated region such that both alleles are silenced. An alternative theoretical mechanism of inactivation of imprinted tumour suppressor genes is through uniparental disomy for the non-expressed parental copy.

Imprinted tumour suppressor genes now implicated in cancer are *TP73*, *NOEY2*, *M6P/IGF2R*, and *H19*, which are discussed below.

1.3.2.2.1 *NOEY2* (*ARH1*)

NOEY2 is an imprinted tumour suppressor gene on chromosome 1p31 identified in a systematic differential display screen for sequences specifically expressed in normal ovarian cells, but not in tumour cells (Yu *et al.* 1999). This gene is a member of the *RAS* family, and may act as a negative growth regulator in cancers of the breast and ovary. Yu *et al.* not only found monoallelic, paternal-specific expression of the gene, but also showed that in four of five patients with LOH at the *NOEY2* locus, the lost allele was the expressed paternal copy, thus leaving only the inactive maternal allele present. In accordance with the Knudson hypothesis described above, the maternal imprint acts as the first hit, and the LOH seen in these cases is the second hit, preferentially removing the only functional copy of the gene. This offers suggestive evidence for a tumour suppressor role for *NOEY2*. Further evidence from this study was the discovery that two of eight breast cancer cell lines were hypermethylated at the *NOEY2* locus, thus resulting in a complete silencing of the gene.

1.3.2.2.2 *IGF2R*

The insulin-like growth factor 2 receptor (*IGF2R*) gene is located at human chromosome 6q26. The protein product encoded by this gene has a threefold function in that it activates the growth inhibitor TGFβ, targets IGF2 for lysosomal degradation, and is involved in the intracellular trafficking of lysosomal enzymes (Hankins *et al.* 1996). Interestingly, the homologous mouse gene is imprinted, being maternally expressed. The underlying basis of the mouse *Igf2r* imprinting, as discussed above, is believed to be the existence of two differentially methylated regions in the gene, one of which may control the expression of an imprinted antisense transcript with a regulatory role (see 1.2.4.1.1). The human gene retains all of the differentially methylated regions that appear to control the monoallelic expression of the mouse homologue, yet, surprisingly still displays biallelic expression (Smrzka *et al.* 1995; Riesewijk *et al.* 1996).

An interesting phenomenon however is that the non-imprinted status of human *IGF2R* gene is not always maintained. A degree of imprinting and maternal-specific

IGF2R expression has been found in a subset of individuals, suggesting the possibility of polymorphic imprinting (Xu *et al.* 1993; Smrzka *et al.* 1995).

Irrespective of any imprinting effect that may exist in humans, *IGF2R* has been proposed as a tumour suppressor gene. Given its functions in degrading IGF2 and activating growth inhibitors, a loss of this gene and these functions would be expected to have some effect in both increasing cell proliferation and reducing apoptosis, which are the hallmarks of tumour development. Along these lines, Hankins *et al.* (Hankins *et al.* 1996) found that 30% of informative breast tumours had LOH; missense mutations of the remaining allele were found in two of five carcinomas *in situ* that they examined, suggesting that allelic loss of *IGF2R* is an early event in breast cancer. Yamada *et al.* 1997 (Yamada *et al.* 1997) performed a similar analysis of patients with dysplastic liver lesions or hepatocellular carcinomas (HCC), and found approximately 60% LOH. Mutations of the remaining allele were detected in 55% of HCCs, suggesting that *IGF2R* acts as a tumour suppressor in liver carcinogenesis. Similarly, Kong *et al.* 2000 (Kong *et al.* 2000) found LOH in 58% of squamous cell lung cancers, with the remaining allele mutated in 6 of the 11 cases analysed.

The overall conclusion from these results is that allelic loss of *IGF2R* occurs frequently in cancer, and its loss would theoretically result in increased levels of IGF2, since IGF2R normally causes degradation of IGF2. In effect, the LOH of *IGF2R* accompanied by mutation of the remaining allele, may be functionally equivalent to loss of imprinting (LOI) of *IGF2*, which results in biallelic expression of the gene, and increased levels of IGF2 protein. Since *IGF2* LOI is known to be a causative factor in many cancers, the LOH of *IGF2R* could feasibly be involved in tumorigenesis as a tumour suppressor gene.

It is apparent therefore, that *IGF2R* acts as a tumour suppressor gene in its own right, and can be eliminated through a two-hit model, with LOH and mutation being the two independent hits. However, the notion of polymorphic imprinting of this gene may also be significant. In accordance with the ideas discussed above, individuals in

whom *IGF2R* is subject to imprinting might be more vulnerable to a complete loss of function of this gene because they have only one active copy and therefore, as in the Knudson hypothesis, require only one additional 'hit' to inactivate *IGF2R*. Thus, these individuals who imprint *IGF2R* may have an increased susceptibility to cancer. To investigate this possibility, Xu *et al.* (Xu *et al.* 1997) looked at Wilms' tumours, in which LOI of *IGF2* is common. They found that in seven of sixteen patients with Wilms' tumour, there was significant repression of the paternally inherited *IGF2R* allele in both the tumour and normal kidney. Approximately 50% of normal fetal kidneys show imprinting of *IGF2R*, suggesting that the partial imprinting of *IGF2R* in a similar proportion of Wilms' tumours is an abnormal persistence of the fetal situation that may contribute to tumour susceptibility.

1.3.2.2.3 *TP73*

The *TP73* gene has been proposed to be an imprinted tumour suppressor gene in neuroblastoma, lung cancer and renal cell carcinoma, yet the evidence for monoallelic expression of this gene is conflicting. However, this subject forms part of the next chapter and will not be discussed here.

1.3.2.2.4 *H19*

The *H19* gene, in the imprinted cluster at 11p15.5, has been proposed as a tumour suppressor based on the preferential maternal LOH of this region in cancers. Additionally, the tumour suppressor activity of the *H19* RNA has been demonstrated in transfection experiments (Hao *et al.* 1993). However, other evidence is contradictory. As discussed above, the LOI of *IGF2* in some cancers is related to the downregulation and usually hypermethylation of *H19*, and this could imply that the loss of *H19* is as important as the gain of *IGF2*. However, there are many examples of LOI and biallelic expression of *H19* in cancer, rather than LOH, and this is strong evidence against a tumour suppressor role. A recent study found inactivation of the *H19* gene by LOH or hypermethylation of the maternally inherited (expressed) allele in seven of eight sporadic hepatoblastomas that were not associated with BWS

(Fukuzawa *et al.* 1999). This supports a tumour suppressor role for *H19*, but studies of this kind are fewer in number than those which have found opposing evidence in the form of biallelic expression of *H19*.

1.4 SUMMARY

Imprinting is a complex phenomenon, which has commanded the focus of a great deal of research in recent years. Attaining a greater understanding of the characteristics and mechanisms of the imprinting process has been a priority, but perhaps of greater significance now is elucidating the relevance of imprinting to disease. Imprinting has a crucial role in normal development, but carries a risk, in that it renders the genome more vulnerable to disruption at imprinted loci. The study of diseases such as Beckwith-Wiedemann syndrome, Prader-Willi syndrome and Angelman syndrome has highlighted some of the consequences of aberrant imprinting, but these diseases may represent only a small proportion of the possible pathological outcomes. A possible imprinting effect has been postulated in many other diseases, including Silver-Russell syndrome and transient neonatal diabetes. Given that estimates of the total number of imprinted genes predict that we may have discovered less than 20% of them to date, and that almost all imprinted genes have been associated with disease (when normal imprinting is disturbed), the key goals must be to identify the remaining imprinted genes, and to characterise them and understand their functions and potential involvement in disease. The widespread and frequent occurrence of aberrant imprinting in cancer illustrates that epigenetic disruption may be of major significance in cancer, and must therefore be a particular focus of future investigation.

CHAPTER 2

MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Chemicals and reagents

2.1.1.1 General chemicals

All chemicals for general use were supplied by Sigma (Sigma-Aldrich Company Ltd) or BDH (Merck Ltd) unless otherwise stated, and were of molecular biology grade. Bacterial media were obtained from Difco U.K. Ltd.

2.1.1.2 Radiochemicals

All radioactive isotopes were supplied by Amersham Life Sciences. These included [α^{32}P]-dCTP at a specific activity of 3000 Ci/mmol for labelling DNA probes, [γ^{32}P]-ATP at a specific activity of 1415 Ci/mmol for labelling oligonucleotide probes, and [α^{32}P]-UTP at a specific activity of 800 Ci/mmol for use in the transcription reaction to generate radiolabelled RNA probes. In addition, for manual radioactive sequencing, the four Redivue^{TM} ^{33}P-labelled dideoxynucleotide terminators were obtained at a specific activity of 1500 Ci/mmol (450μCi/ml).

2.1.1.3 Enzymes

Restriction enzymes were obtained from New England Biolabs. Of particular relevance to the work in this study is that a subset of the restriction enzymes used are sensitive to CpG methylation, such that cleavage is blocked when CpG sites within the recognition sequence are methylated. Exonuclease I for purification of PCR products prior to sequencing was supplied from Amersham as part of a kit also containing shrimp alkaline phosphatase. The Klenow fragment of *E.coli* DNA polymerase I was provided by Boehringer. *Taq* polymerase (5U/μl) and T4 DNA ligase were supplied by Gibco BRL Life Technologies. The SP6 and T7 RNA polymerases were supplied by Ambion. Desiccated proteinase K and RNase A were

obtained from Sigma. Superscript™ II reverse transcriptase was supplied by Gibco BRL Life Technologies. The RNase inhibitor RNAguard™, derived from human placenta, was supplied by Amersham Pharmacia at a concentration of 29700 units/ml.

2.1.1.4 Nucleic acid markers

The 1kb ladder and 100bp ladder DNA markers were obtained from Gibco BRL Life Technologies. For ABI gels, the Perkin Elmer GENESCAN-350 TAMRA marker was used. For pulsed field gel electrophoresis the Mid Range PFG Marker 1 from New England Biolabs was used.

2.1.1.5 Components of PCR reactions

Taq polymerase is detailed elsewhere. The PCR buffer and $MgCl_2$ solution supplied with the *Taq* polymerase were generally used. Stocks of dNTPs were obtained from Gibco BRL Life Technologies, at a concentration of 100mM each, and a working stock consisting of 25mM each dNTP was prepared. PCR primers were synthesised at the 50nmol scale by either Genosys, or Life Technologies. A working stock was prepared by reconstituting each primer in distilled H_2O to a concentration of 100pmol/µl, and then using a subsequent dilution of 20pmol/µl for PCR. The sequences of the primers used throughout this thesis are given in Appendix 2.A.2. They are referred to by name at the appropriate points in the text of each relevant chapter.

2.1.2 DNA samples

All DNA samples were stored at 5°C.

2.1.2.1 Normal DNA controls

A set of normal control human genomic DNAs was used frequently for optimising PCRs, generating probes, and other general uses. These had previously been prepared from blood extractions, and the stocks were generally at a concentration of 0.5µg/µl. Working dilutions of 100ng/µl were prepared.

2.1.2.2 Parthenogenetic DNA

A source of human parthenogenetic DNA was available from the chimaeric patient described by Strain *et al.* (Strain *et al.* 1995). This patient has chimaerism between normal cells with biparental contribution, and parthenogenetic cells. The leukocytes are completely parthenogenetic, and DNA from a lymphoblastoid cell line established from these cells was used for analysis of imprinting in this study.

2.1.2.3 Fetal DNA panel

DNA and RNA samples from a panel of 66 first trimester fetuses had previously been collected and described (Campbell *et al.* 1994). The panel comprises fetal and corresponding maternal DNAs, together with fetal RNAs from a range of tissues, particularly muscle, eye, stomach, testis, cord, adrenal, brain, liver, kidney, heart, lung and gut. Table 1 shows further details of the panel.

Table 1. Details of the panel of first trimester fetal DNAs.

NUMBER	CODE	FETAL DNA AVAILABLE?	MATERNAL DNA AVAILABLE?	NUMBER OF FETAL RNAS AVAILABLE
1'	MB040267	Yes	Yes	5
2'	LH260874	Yes	Yes	8
3'	MW080265	Yes	Yes	3
4'	LF010174	Yes	Yes	5
6'	MM060567	Yes	Yes	2
7'	CA020275	Yes	Yes	9
8'	LH200770	Yes	Yes	4
9'	MC260166	Yes	Yes	6
10'	JB040570	Yes	Yes	2
11'	AF281169	Yes	No	6
12'	LM250774	Yes	Yes	4
13'	HM170958	Yes	Yes	3
14'	AM121062	Yes	Yes	7
15'	AM300969	Yes	Yes	7
16'	PM26	Yes	Yes	1
17'	MC151165	Yes	Yes	7
18'	MD311268	Yes	Yes	3
19'	DMc090964	Yes	Yes	6
20'	TS070870	Yes	Yes	6
21'	FB111171	Yes	Yes	8
22'	JC290868	Yes	Yes	5
23'	HW300462	Yes	Yes	4
24'	LR190174	Yes	Yes	10
25'	MT150964	Yes	Yes	2
26'	RL	Yes	Yes	3
27'	SS130773	Yes	Yes	8
28'	JS020673	Yes	Yes	6
29'	SG281070	Yes	No	5
30'	SH2401	Yes	Yes	7
31'	LR0906	Yes	Yes	6
A	SN301070	Yes	Yes	3
B	CS280766	Yes	Yes	5
C	AS020172	Yes	No	8

D	CB230357	Yes	No	7
E	SG70	Yes	Yes	5
F	AJ180558	Yes	No	6
G	NG100369	Yes	Yes	1
H	RT281475	Yes	Yes	1
I	HA040275	Yes	Yes	8
J	JM230372	Yes	Yes	4
K	JA161159	Yes	No	2
L	JR170474	Yes	Yes	6
M	FC070271	Yes	Yes	3
N	JH210566	Yes	Yes	8
O	AJ180558	Yes	Yes	6
P	TS070870	Yes	No	6
Q	HW150273	Yes	No	6
R	CA250568	Yes	Yes	2
S	CT111275	Yes	Yes	5
T	LB020170	Yes	Yes	6
U	AS080172	Yes	Yes	8
V	JF101272	Yes	Yes	7
W	JC230671	Yes	Yes	4
X	GK011268	Yes	Yes	2
1	EH131162	Yes	Yes	3
2	JA161159	Yes	No	2
3	HW150273	Yes	No	6
5	DL040371	Yes	Yes	7
6	AC150452	Yes	Yes	7
7	DB200170	Yes	No	3
8	DH270971	Yes	Yes	8
9	DB280773	Yes	Yes	3
10	MS100972	Yes	Yes	8
11	LM120370	Yes	Yes	7
13	CM311273	Yes	Yes	1
14	CM181169	Yes	Yes	7

2.1.2.4 Neuroblastoma DNAs

A panel of 42 neuroblastoma cell lines and tumours, Merkel cell carcinoma cell lines and tumours, and schwannomas was provided by Dr. F. Speleman. Full details of these are given in Chapter 3.

2.1.2.5 Breast cancer and bladder cancer DNAs

A panel of 11 pairs of breast tumour and matched normal DNA, and 8 pairs of bladder tumour and matched normal DNA was provided by Dr. A. McCann, University College Dublin.

2.1.2.6 Breast cancer cell lines

Six breast cancer cell lines were provided by Dr. V. Speirs, Molecular Medicine Unit, St. James's Hospital, Leeds. These were MDA-MB-435, MDA-MB468, T47D, ADR-MCF7, MCF7, and MCF7-Clone 9. The cell lines were grown under standard cell culture methods, and DNA extracted as described in 2.2.1.3 below.

2.1.2.7 Lymphoma samples

A panel of 21 B-cell non-Hodgkin's lymphomas was provided by Professor K. MacLennan. These were provided as blocks of excised tissue, frozen at -70°C. DNA and RNA were extracted as described in 2.2.1.3 and 2.3.1.

2.1.3 Microbiological resources

2.1.3.1 Libraries

2.1.3.1.1 Gridded filters of PAC library

The human PAC library RPCI1, constructed by Pieter de Jong and his group at the Roswell Park Cancer Institute, Buffalo, was obtained through the UK HGMP

Resource Centre in the high density gridded filter format. This comprises seven double spotted 22.2x22.2 cm filters for screening by hybridisation.

2.1.3.1.2 Gridded filters of chromosome 1 cosmid library

The gridded filters of the LL01NC01 chromosome 1 cosmid library constructed at Lawrence Livermore National Laboratory (LLNL) (Trask *et al.* 1991), were obtained through the UK HGMP Resource Centre.

2.1.3.1.3 Gridded filters of CEPH Mega-YAC library

The CEPH "mega" YAC library, constructed at CEPH (Centre d'Etudes du Polymorphisme Humain) (Chumakov *et al.* 1992) was obtained through the UK HGMP Resource Centre as 8 double spotted 22.2 x 22.2 cm high density gridded filters for screening by hybridisation.

2.1.3.2 Clones

PAC, YAC, and cosmid clones from the libraries mentioned above were supplied by the UK HGMP Resource Centre. In addition, I.M.A.G.E. clones were obtained from the I.M.A.G.E. consortium through the UK HGMP Resource Centre.

2.1.3.3 Vectors

The vector pCR™II was supplied by Invitrogen. The vector TVEC was supplied by Promega. The pUC18 plasmid vector (Yanisch-Perron *et al.* 1985) was obtained from Stratagene.

2.1.3.4 Electrocompetent cells

For transformations, the *E. coli* strain JM109 (Epicurean Coli®) was obtained from Stratagene.

2.1.3.5 Maxiprep and miniprep kits

The Qiagen maxi-prep kit was used for large-scale DNA preparations from PAC clones. Minipreps were either performed using the alkaline lysis method, or the SNAP™ miniprep kit.

2.1.4 Mapping resources

2.1.4.1 Somatic Cell Hybrid Panel

For PCR mapping to individual chromosomes, the human monochromosomal somatic cell hybrid DNA panel (Originators: S. Povey and N. K. Spurr) (Kelsell *et al.* 1995) was obtained from the UK HGMP Resource Centre.

2.1.4.2 Radiation Hybrid Mapping panel

For finer PCR mapping to known markers, the Genebridge 4 radiation hybrid DNA panel (Originators: Peter Goodfellow and Jean Weissenbach) (Gyapay *et al.* 1996) was obtained through the UK HGMP Resource Centre.

The resulting data was analysed at RhyME (Radiation Hybrid Mapping Environment), and the Sanger Centre.

2.1.5 Solutions and buffers

All solutions and buffers were prepared using distilled water, and autoclaved where appropriate under the conditions 15 psi, 121°C for 30 minutes. Details of the components of general solutions and buffers are given in Appendix 2.A.2.

2.1.6 Bioinformatics

The following software and bioinformatics web sites were used during this study:

Genescan™ software by Applied Biosystems

Pubmed/Entrez databases at www.ncbi.nlm.nih.gov

GDB database at www.gdb.org

Sanger Centre at www.sanger.ac.uk

LDB genetic map at http\\cedar.genetics.soton.ac.uk

Primer3 software at: www-genome.wi.mit.edu/cgi-bin/primer/primer3

GCG-HGMP-RC at www.hgmp.mrc.ac.uk/Registered/Option/gcg.html

NIX-HGMP-RC at www.hgmp.mrc.ac.uk/Registered/Webapp/nix/

2.2 DNA METHODS

2.2.1 DNA extractions

2.2.1.1 DNA preparation from PAC clones

DNA was extracted from PAC clones using the Qiagen® plasmid maxi kit. The basis of the protocol is an alkaline lysis procedure, followed by binding of the plasmid DNA to a Qiagen® tip column containing the Qiagen® anion exchange resin. RNA, protein and low molecular weight impurities are selectively removed by a medium salt wash, and the plasmid DNA is eluted using a high salt buffer, then precipitated with isopropanol. Buffers and solutions mentioned below were provided in the kit. Centrifugations were carried out in a Sorvall centrifuge, using the GSA rotor (14.5cm radius) for the initial pelleting of the cells, and the SS-34 rotor (10.7cm radius) for all subsequent centrifugation steps. A 500ml culture of a single streaked colony was grown overnight at 37°C with shaking at 225rpm. Following centrifugation at 6000×g for 15 minutes at 4°C, the pellet was resuspended in 10ml of buffer P1 (50mM Tris-HCl pH 8.0, 10mM EDTA, 100µg/ml RNase A). To lyse

the cells, 10ml of buffer P2 (200mM NaOH, 1% SDS) was added, and the tubes inverted 4-6 times to mix. Precipitation was performed by adding 10ml of chilled buffer P3 (3M potassium acetate, pH 5.5), mixing gently, and incubating on ice for 20 minutes. After centrifugation at 20000×g for 30 minutes at 4°C, the supernatant was removed, and centrifuged again for 15 minutes. A Qiagen tip was equilibrated by allowing 10ml buffer QBT (750mM NaCl, 50mM MOPS pH 7.0, 15% isopropanol) to run through by gravity flow, and then the supernatant was applied to the tip and allowed to enter the resin. The tip was washed with two 30ml washes of buffer QC (1M NaCl, 50mM MOPS pH 7.0, 15% isopropanol), then the DNA was eluted with 15ml buffer QF (1.25M NaCl, 50mM Tris-HCl pH 8.5, 15% isopropanol). The DNA was precipitated from the eluate by adding 10.5ml isopropanol, mixing, and centrifuging at 15000×g for 30 minutes at 4°C. The pellet was washed with 5 ml 70% ethanol, then air dried and redissolved in 200µl TE, pH 8.0.

2.2.1.2 Miniprep DNA extractions

2.2.1.2.1 Alkaline lysis method (Modification of the method of Ish-Horowicz and Burke (Ish-Horowicz, Burke 1981))

The solutions used in this method are detailed in Appendix 2.A.2. Centrifugations were performed in an Eppendorf 5415C benchtop microfuge, using a rotor of radius 8.2cm.

A sterile tube containing 4ml of LB with 20µg/ml ampicillin was inoculated with a single colony and incubated at 37°C overnight with shaking (~225rpm). Of the 4ml, 1ml was taken for a glycerol stock, and the remaining 3ml were centrifuged at 4000rpm to pellet the cells. The pellet was resuspended in 100µl solution I by vigorous vortexing followed by incubation at room temperature for 5 minutes. 200µl of solution II was added, the mixture shaken, and incubated on ice for 5 minutes. 150µl of ice cold solution III was added, the mixture shaken, and incubated on ice for 5 minutes. Tubes were then spun at 13000rpm at 4°C for 3 minutes. The

supernatant was transferred to a tube containing 450µl phenol:chloroform:isoamyl alcohol (25:24:1), vortexed, spun for 1 minute, and the aqueous phase removed to a fresh tube. 900µl of 100% ethanol was added and precipitation allowed to proceed for 5 minutes at room temperature, followed by spinning at 13000 rpm at room temperature for 5 minutes. The pellet was washed with 70% ethanol, air-dried, and resuspended in 50µl TE containing 50µg/ml RNase.

2.2.1.2.2 SNAP™ miniprep kit

High quality miniprep DNA for use in manipulations such as sequencing was prepared with the SNAP™ (Simple Nucleic Acid Prep) kit from Invitrogen. The kit provides Resuspension Buffer, Lysis Buffer, Precipitation Salt, Binding Buffer, Wash Buffer, and 1×Final Wash.

A 4ml culture of LB containing 20µg/ml ampicillin was inoculated with a single colony and incubated at 37°C overnight with shaking (~225rpm). Of the 4ml, 1ml was taken for a glycerol stock, and the remaining 3ml were centrifuged at 4000rpm to pellet the cells, and the pellet resuspended in 150µl Resuspension Buffer. The cells were lysed with 150µl Lysis Buffer, and incubated at room temperature for 3 minutes. 150µl ice cold Precipitation Salt was added, then the contents of the tubes mixed, and centrifuged at 14000×g for 5 minutes. The supernatant was removed to a sterile tube, 600µl Binding Buffer added, the solution mixed, and transferred onto a SNAP™ miniprep column/collection tube. The column/tube was then centrifuged at 3000×g for 30 seconds, to bind the DNA to the column. To wash the DNA, 500µl of Wash Buffer was added to the column, and centrifuged as in the previous step. A final wash was performed in the same way with 900µl 1× Final Wash, after which the column/collection tube was centrifuged at 13000rpm at room temperature for 1 minute to dry the resin. The DNA was eluted into 60µl TE, which was added to the column, incubated for 3 minutes at room temperature, and centrifuged at 13000rpm for 30 seconds. To check the yield and quality of the DNA, 2-3µl was electrophoresed on a 1% agarose gel alongside samples of known concentration.

2.2.1.3 Extraction of DNA from lymphomas and breast cancer cell lines

The following protocol was used to extract DNA from both cells and tissues. Tissues were minced with a scalpel, on ice, before adding the digestion buffer (details of this buffer are given in Appendix 2.A.2). Cells were spun down to a pellet.

To each tube containing a cell pellet or minced tumour tissue, 500µl digestion buffer and 30µl proteinase K (10mg/ml) were added, and the tubes vortexed vigorously. The digestion buffer contained 50mM Tris-HCl (pH 8.0), 100mM NaCl, 5mM EDTA and 1% SDS in dH_2O. The tubes were incubated overnight at 50°C in a water bath, after which 500µl phenol:chloroform:isoamyl alcohol (25:24:1) was added to each tube. The tubes were vortexed, and centrifuged at 13,500rpm for 5 minutes at room temperature. The supernatant was removed to a fresh tube, 500µl phenol:chloroform:isoamyl alcohol (25:24:1) added, the tubes vortexed, and centrifuged again at 13,500rpm for 5 minutes at room temperature. The supernatant was transferred to a fresh tube, and 20µl 5M NaCl and 520µl 100% ethanol were added. The contents of the tube were mixed, and then the tubes incubated at -80°C for 30 minutes to precipitate the DNA. The tubes were centrifuged at 13,500rpm for 10 minutes at room temperature, and the pellet washed in 70% ethanol, then allowed to air dry for 5-10 minutes. The pellet was resuspended in 50µl TE buffer, pH 8.0, and then subjected to an RNase treatment at 37°C for 1 hour, with 1µg/ml RNase. To check the quality and estimate the yield of the DNA, 5µl of the sample was run on a 1% agarose gel alongside samples of known concentration, and visualised with ultraviolet light.

2.2.2 Polymerase Chain Reaction (PCR)

All PCR reactions were carried out in a total volume of 50µl. The reaction mix contained:

1× PCR buffer (10× stock contains 15mM $MgCl_2$, 100mM Tris-HCl (pH 8.0), 500mM KCl)

10pmol forward primer
10pmol reverse primer
200µM each dNTP
1 unit *Taq* polymerase
Template DNA (usually 100ng genomic DNA)
Distilled H_2O up to 50µl

Cycling conditions were usually 30 to 40 cycles of denaturation at 94°C for 1 minute, annealing at X°C for 1 minute, and extension at 72°C for 2 minutes.
Conditions varied slightly between each PCR, depending on the annealing temperature required, the expected product length, and other variable factors. In some cases where dimerisation of the two primers was judged to be problematic, a hot start was used. In cases where a PCR failed repeatedly, the addition of higher concentrations of $MgCl_2$, or 10% dimethylsulphoxide (DMSO), or 0.1%-1% Triton X-100 was often found to assist in the generation of a product.

2.2.3 Gel electrophoresis

2.2.3.1 Agarose gel electrophoresis

For the purpose of analysing PCR products, checking DNA quality/concentration, and analysing digestions, electrophoresis was usually performed in 1% TAE gels, prepared with 1g agarose per 100ml 1×TAE buffer, and containing 0.5µg/ml ethidium bromide. Gels were run in 1×TAE buffer, at 5 Volts per cm, for 30-60 minutes, depending on the size of the fragments of interest. Gels were visualised using an ultra violet trans-illuminator.

For resolving smaller fragments, 2% or 3% gels made with NuSieve™ agarose were used.

2.2.3.2 Pulsed field gel electrophoresis

For the separation and resolution of larger DNA fragments, such as the products of digestion of PAC clones with rare-cutting enzymes, pulsed field gel electrophoresis was employed, using the contour-clamped homogeneous electric field (CHEF-DRII) system from Biorad. This system comprises a hexagonal arrangement of 24 electrodes, and operates on the basis of alternating the electric field between pairs of electrodes that are orientated at 120° with respect to each other. Gels were 1% agarose (14×13cm), and run in 0.5×TBE at 6V/cm, 14°C. The switch times between the alternating electric fields were ramped from 1 to 10 seconds during a 24 hour period. Gels were stained after electrophoresis by immersing in 0.5×TBE containing 0.5μg/ml ethidium bromide, for 30-60 minutes, and then visualised with ultraviolet light.

2.2.3.3 Polyacrylamide gel electrophoresis

2.2.3.3.1 Non-denaturing polyacrylamide gels

In cases where the resolution of similar sized fragments of DNA was required, e.g. restriction digests, polyacrylamide gels were used. These were usually 20cm×20cm×0.1cm, 6% acrylamide, 0.5×TBE, according to the recipe in 2.A.2. Biorad gel rigs were used. Glass plates were cleaned with ethanol before use. Gels were poured and allowed to polymerise for 1 hour before use.

2.2.3.3.2 Denaturing polyacrylamide gels

For the separation of single stranded DNA, e.g. transcription products, denaturing 5% acrylamide, 0.5×TBE, 8M urea gels were used. The recipe is given in 2.A.2. Gels were poured and run as for non-denaturing gels.

2.2.4 Digestion of DNA with restriction endonucleases

2.2.4.1 Digestion of human genomic DNA

For the purposes of Southern blotting, 10µg (usually 20µl) of human genomic DNA was digested overnight with 30-50 units of enzyme (up to a maximum of 10% of the reaction volume), in a total volume of 50µl. Incubations were performed in water baths or heating blocks at the stated optimum temperature for each enzyme.

2.2.4.2 Digestion of PCR products

Restriction enzyme digestions of PCR products were typically performed in volumes of 20µl, comprising 10µl PCR product, 2µl appropriate buffer, 2µl 10× BSA (if required), 10-50 units of enzyme (up to a maximum of 10% of the reaction volume, i.e. 2µl here), and dH_2O to 20µl. The reactions were then incubated at the stated optimum temperature, in a water bath or heating block, for 4-16 hours.

2.2.4.3 Digestion of PAC clones

For the purposes of analysing the restriction sites within a PAC clone insert sequence, or for subcloning of a PAC clone, digestion of 1µg DNA was performed in 100µl reaction volumes containing up to 50 units of enzyme. Double digestions, where required, were performed in the buffer recommended by New England Biolabs. In cases where the combination of enzymes was not complementary to any of the available buffers, sequential digests were performed by digesting first with one enzyme, then performing a phenol chloroform extraction and ethanol precipitation, and subsequently digesting again with the second enzyme.

2.2.5 Southern blotting

Southern blotting was used to transfer DNA from agarose gels to Nylon membrane according to a modification of the original method of Southern, (Southern 1975).

After electrophoresis the gel was immersed in 0.25M HCl for 15 minutes with gentle shaking, then rinsed in dH_2O. The transfer was performed in a glass dish, containing a transfer platform and a 3MM filter paper wick, which was in contact with a reservoir of 0.4M NaOH. The gel was placed on the filter paper wick, and a gel-sized piece of Hybond-N^+ nylon membrane (from Amersham) placed on the surface of the gel. Three pieces of 3MM filter paper were placed on top of this, followed by a stack of absorbent paper towels, and a weight. Cling film was used to cover the area around the gel, to prevent evaporation of the solution. The transfer was allowed to proceed, by capillary action, for 16 hours. After transfer, the membrane was rinsed in 2×SSC for 5 minutes, and then baked between two sheets of filter paper for 2 hours at 80°C.

2.2.6 Labelling probes for hybridisation

2.2.6.1 Oligonucleotide probes

Oligonucleotides of 18-25bp were labelled with [γ^{32}P] ATP using the enzyme T4 polynucleotide kinase (PNK). Labelling reactions of 10µl volume were set up containing 30-50ng of the oligonucleotide (usually 1µl of a 5pmol/µl stock), 1µl 10× PNK buffer, 0.5µl (5 units) PNK, 3µl [γ^{32}P] ATP, and 4.5µl dH_2O. Reactions were incubated at 37°C for 30 minutes.

2.2.6.2 PCR product probes

Probes were labelled by the random priming method (Feinberg, Vogelstein 1983), either by a standard protocol, or by using the Pharmacia kit method, both of which are described below.

2.2.6.2.1 Standard method

Typically, 50ng of DNA was labelled. The DNA solution was increased to a volume of 33µl with the addition of dH_2O, denatured by incubating in a boiling water bath for 10 minutes, and then placed on ice to prevent renaturation. To the denatured DNA was added 12µl OLB (containing random 6mer primers and a buffered solution of cold dATP, dGTP, and dTTP), 4µl [α^{32}P]-dCTP and 1µl Klenow enzyme. The labelling reaction was allowed to proceed at room temperature for 4 hours, after which the probe was denatured again, and then added to the hybridisation reaction.

2.2.6.2.2 ^{T7}QuickPrime® kit method

The labelling of PCR product probes, or digested fragments of plasmids, was performed using the ^{T7}QuickPrime® kit from Pharmacia Biotech. The DNA of interest was first electrophoresed in a low melting point agarose gel, and then the band (containing at least 250ng of DNA) was excised from the gel and weighed. Distilled water was added to a volume of 3ml per gram of gel slice, then the mixture heated at 65°C to melt the agarose. For the labelling reaction, a maximum of 25µl of this mixture was used, following preparation by heating at 95°C for 7 minutes, and then incubating at 37°C for 10 minutes.

Labelling reactions were set up comprising the following reagents:
25µl denatured DNA (25-50ng)
10µl Reagent Mix (provided in the kit)
5µl [α-^{32}P] dCTP (3000 Ci/mmol)
Distilled H_2O to a final volume of 49µl

1µl of T7 DNA polymerase (provided in the kit) was added, then the contents of the tube mixed gently, and centrifuged briefly, before being incubated at 37°C for 15 minutes. Immediately before hybridisation, the labelled probe was denatured at 95° for 3 minutes.

2.2.7 Hybridisation of probes to PAC filters/Southern blots

2.2.7.1 PCR product probes – standard method

Filters were immersed in 2×SSC buffer, then rolled up and placed in hybridisation bottles, each filter being separated from the others with nylon membranes. 20ml prehybridisation buffer that had been preheated to 65°C was added to the bottles along with 200µl (10mg/ml stock) sonicated salmon sperm DNA (denatured by incubating for 10 minutes in a boiling water bath), and the filters were then allowed to prehybridise for 1 hour, rotating, in an oven at 65°C.

The prehybridisation mix was poured away, and 10ml of pre-heated hybridisation mix was added to the bottles. The labelled probe and 100µl salmon sperm DNA (10mg/ml stock) were denatured by incubating in a boiling water bath for 10 minutes, and were then added to the bottles. The hybridisation was allowed to proceed overnight, at 65°C.

After the hybridisation the filters were subjected to two 20-minute washes in 2×SSC/0.1% SDS, and one 20-minute wash in 1×SSC/0.1% SDS, all at 65°C. The filters were wrapped in cling film, and exposed to autoradiographic film with intensifying screens at -70°C, for between 4 and 24 hours.

2.2.7.2 Oligo Probes – Quick Oligo Hyb method

Filters were soaked in 2×SSC for 5 minutes, and then rolled up and placed in the hybridisation bottles. 10ml of Quick Oligo Hyb Mix was heated up to 68°, then added to the bottles along with 100µl denatured salmon sperm DNA (10mg/ml stock). Prehybridisation of the filters was allowed to continue for 1 hour at 68°C, and then the labelled oligo probe was added to each bottle, and hybridised overnight. Three to four 5-minute washes of the filters were performed with 4×SSC/0.1% SDS

at 50°C, and the filters were wrapped in cling film and exposed to autoradiographic film in light-proof cassettes with intensifying screens at -70°C for 4-24 hours.

2.2.8 PAC subcloning

For the subcloning of *Bam*HI and *Bgl*II fragments of PAC clones, the vector pUC18 was used. Prior to ligation with the digest fragments, the vector was digested with *Bam*HI itself, and the phosphate groups removed by treatment with shrimp alkaline phosphatase, to prevent religation of the vector to itself. For subcloning of *Not*I-*Sac*I fragments, the vector pCR™II was used, following digestion with *Not*I and *Sac*I; for subcloning of *Asc*I-*Sac*I fragments the vector TVEC was used, following digestion with *Mlu*I and *Sac*I.

2.2.8.1 Ligation

Approximately 1µg of PAC DNA was digested with at least 50 units of the appropriate enzymes depending on the fragments required to be subcloned, e.g. *Not*I-*Sac*I double digest, *Asc*I-*Sac*I double digest, *Bam*HI or *Bgl*II single digests. The fragments were recovered by phenol-chloroform extraction and ethanol precipitation. Ligation reactions were set up containing 40ng of the appropriately digested vector, 500ng of the insert (digested PAC clone DNA), 3µl 5×ligase buffer, and 1µl of T4 DNA ligase in a total volume of 15µl, made up with distilled H_2O. A control ligation reaction was set up in each case, comprising the same components but distilled H_2O in place of the insert. Ligation reactions were incubated at room temperature overnight, followed by ethanol precipitation. The pellet was resuspended in 5µl H_2O.

2.2.8.2 Transformation by electroporation

The Gene Pulser™ (Biorad) apparatus was set to 25µF, 2.5kV and the controller to 200Ω. On ice, 1µl of the precipitated ligation reaction was added to 40µl thawed

electrocompetent cells (see below for preparation of these cells), then the mixture transferred into a chilled electroporation cuvette. The cuvette was pulsed according to the conditions above, and then 1ml LB broth was immediately added to the cuvette. The contents of the cuvette were transferred to an autoclaved Eppendorf tube, and incubated for 1 hour at 37°C with shaking at 225rpm, then 100µl and 200µl were plated out onto fresh X-gal/IPTG/ampicillin plates (see 2.A.2), and incubated overnight.

2.2.9 Preparation of electrocompetent cells

A stock of *E. coli* strain JM109 (Epicurean Coli®) cells was obtained from Stratagene. The following protocol was used for preparation of the cells for transformation.

One litre of LB was inoculated with 1/100 volume of an overnight culture from a freshly streaked JM109 colony. The cells were grown at 37°C with vigorous shaking (225rpm) until they reached early to mid-log phase, defined by an OD_{600} of 0.5 to 0.7.

The culture was transferred into 50ml Falcon tubes, and placed on ice for 15-30 minutes. The tubes were centrifuged at 4000rpm for 5 minutes at 4°C. The supernatant was removed, and the pellets resuspended in 50ml ice-cold water. The tubes were centrifuged again, and the pellets resuspended in half the volume of ice-cold water, combining two tubes into one. The tubes were centrifuged again, and the pellets resuspended in 20ml ice-cold water, combining all tubes into one. A final centrifugation was performed, and the pellet resuspended in 3ml 10% ice-cold glycerol. The suspension was divided into aliquots of 45µl, which were immediately frozen in liquid nitrogen, and then stored at -70°C for future use.

2.2.10 PCR product cloning

Cloning of PCR products was performed using the pGEM®-T Easy Vector kit from Promega, according to the manufacturer's protocol, as follows:

2.2.10.1 Ligation

Ligation reactions were set up at room temperature according to the manufacturer's protocol, and contained 5µl Rapid Ligation Buffer, 1µl (50ng) pGEM®-T Easy Vector, 1µl PCR product, 1µl (3 Weiss units/µl) T4 DNA ligase, and deionized water to a total volume of 10µl. (NB. The buffer, ligase and vector were components of the kit). The PCR product insert was always used directly from a 50µl reaction without any purification step. After mixing, the ligation reactions were incubated at room temperature for 4 hours. A control reaction was set up each time containing dH_2O as a replacement for the insert DNA.

2.2.10.2 Transformation by heat shocking

On ice, 2µl of each ligation was added to a sterile 1.5ml microcentrifuge tube. The frozen JM109 competent cells were placed in an ice bath until just thawed, then they were gently mixed, and 50µl carefully transferred to each ligation tube prepared as above. The tubes were placed on ice for 20 minutes, and then heat shocked in a water bath at exactly 42°C for 45-50 seconds. The tubes were returned to the ice for 2 minutes, and then 950µl LB medium was added, and the tubes incubated for 1 hour at 37°C with shaking (~150rpm).

100µl of each transformation culture was plated out onto duplicate LB/amp/IPTG/X-gal plates (for recipe see 2.A.2), and the plates incubated overnight for 16 hours at 37°C. Each plate of the reaction containing PCR product insert typically yielded 100-200 white colonies, which generally contained inserts of the expected size. The control plates should not yield any white colonies, but may yield a background level of blue colonies (no inserts), as a result of non T-tailed or undigested vector.

To confirm the presence of the desired insert, colony PCR was performed with vector primers M13Forward and M13Reverse, by making up a PCR mix, then transferring a small amount of the colony from the end of a Gilson tip into the mix, and proceeding as for a normal PCR, except for allowing an initial 10 minute incubation at 94°C to help lyse the cells. DNA was prepared from the positive colonies by the SNAP miniprep kit, as described elsewhere.

2.2.11 Purification of PCR products for sequencing and other subsequent manipulations

2.2.11.1 Shrimp alkaline phosphatase/exonuclease I purification

To remove single stranded primers and dNTPs from the PCR reaction prior to sequencing, 1µl (1 unit) SAP (shrimp alkaline phosphate) and 1µl (10 units) exonuclease I were added to 5µl PCR product (from a 50µl reaction), and incubated at 37°C for 15 minutes. This was followed by incubation at 80°C for 15 minutes, to inactivate the enzymes. After this treatment, 1-3µl of the product was used directly in the thermosequenase radiolabelled terminator cycle sequencing reaction.

2.2.11.2 Geneclean purification kit

For purification of DNA from agarose, the GENECLEAN kit from Bio101 was used according to the manufacturer's protocol, as follows:

The band of interest was excised from the gel with a scalpel, taking appropriate precautions to avoid personal UV exposure. The band was weighed, and then placed in a tube with three volumes of 6M sodium iodide, and the tube incubated at 50°C for 5 minutes until the agarose had dissolved. Glassmilk, an aqueous suspension of proprietary silica matrix, was added – the volume was dependent on both the amount of DNA expected in the gel slice, and the total volume of the sodium iodide solution

at this stage, but was generally between 5 and 20µl. After adding the glassmilk the tubes were incubated for approximately 15 minutes on a mechanical shaker to allow the DNA to bind to the silica matrix. The glassmilk was then pelleted by spinning at 13000rpm for 5 seconds. The glassmilk (and bound DNA) was washed three times by resuspending the pellet in 500µl NEW Wash, spinning for 5 seconds, and discarding the supernatant. Following the removal of the supernatant from the third wash, the tubes were re-spun, in order to remove any residual liquid, before drying the pellet for 5-10 minutes at room temperature. To elute the DNA from the glassmilk, the pellet was resuspended in a volume of water equal to the amount of glassmilk originally added. After a 30 second spin, the supernatant now containing the DNA was removed to a fresh tube.

Sodium iodide (6M), glassmilk and NEW Wash (containing NaCl, Tris, EDTA and 50% ethanol) were provided in the GENECLEAN kit.

2.2.12 ^{33}P radiolabelled terminator cycle sequencing

Templates for sequencing by this method were either PCR products that had been purified by the SAP/exonucleaseI method described above, or plasmid DNA or PAC DNA which were used directly in the reaction. The kit used for this method was the thermosequenase radiolabelled terminator cycle sequencing kit from Amersham.

2.2.12.1 Sequencing reactions

The termination mixes were prepared for each [α^{33}P] labelled ddNTP (ddATP, ddCTP, ddGTP and ddTTP) by combining 1µl of dGTP termination master mix (containing 7.5µM each of dATP, dCTP, dGTP and dTTP) and 0.25µl of each radioactively labelled ddNTP. In another tube, a reaction mixture was prepared comprising: 1µl of 10× reaction buffer (260mM Tris-HCl pH 9.5, 65mM $MgCl_2$), 1-7µl of template DNA (amount depending on the source and nature of the template e.g. plasmid, PCR product), 1µl of sequencing primer (at a concentration of 3µM), 1µl (4U) of Thermosequenase DNA polymerase, and sterile H_2O to a total volume of

10µl. To each of the prepared termination mix tubes (A, C, G, and T), 2.25µl of this reaction mixture was added, and the mixture overlaid with mineral oil.

The cycle sequencing reactions were performed in a Hybaid Omnigene Thermal Cycler, in 96 well microtitre plates. Where the standard dGTP mastermix was used, the cycling conditions were 40-50 cycles of 95°C for 30 seconds, 55°C for 30 seconds, and 72°C for 1 minute. In the cases where the dITP mastermix was used to reduce the effect of GC compressions on the gel (the dITP mastermix contains 7.5µM each of dATP, dCTP, dTTP and 37.5µM dITP), the conditions were 50 cycles of 95°C for 30 seconds, 50°C for 30 seconds, and 60°C for 7 minutes. On completion of the cycling reaction, 2µl of stop solution (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF) was added to each termination reaction. The samples were then denatured at 95°C for 3 minutes and 2.5µl of each was loaded immediately onto a 6% glycerol tolerant sequencing gel, as described below.

2.2.12.2 Preparation and running of acrylamide sequencing gels

The products of cycle sequencing reactions were separated by electrophoresis on 0.4mm×38cm×50cm Biorad sequencing gel rigs. The acrylamide gel mix was prepared as a 950ml stock (to be made up to the equivalent of 1 litre with 20× glycerol tolerant buffer) comprising 8M urea, 6% acrylamide/*N,N'*-methylenebisacrylamide (19: 1, Acrylogel mix 5- Sigma), and distilled water. Each gel was then prepared with 95ml of this stock and 5ml 20× glycerol tolerant buffer, and polymerised by adding 150µl 25% ammonium persulphate and 100µl NNN'N' tetramethylenediamine (TEMED) immediately before pouring. Gels were allowed to polymerise for 2 hours at room temperature, before running in 1× glycerol tolerant buffer at 90W for 2-6 hours, depending on the length of readable sequence required.

2.2.12.3 Visualisation of results

Following the electrophoresis of sequencing products, the gel was transferred to 3MM filter paper, covered in Saran wrap, and dried under vacuum at 80°C for 1 hour

in a Biorad gel drier, after which the Saran wrap was removed. The gel was then exposed to autoradiographic film in a light-proof cassette at room temperature, for 1-2 days, and developed.

2.2.13 ABI analysis of polymorphic microsatellite markers

Five polymorphic dinucleotide repeat microsatellite markers from the Cedar map (GDB) were used for detection of loss of heterozygosity (LOH) at chromosome 6q24-25. In addition, three novel dinucleotide repeat microsatellites were identified within the PAC clone sequences of 340-H11, 197-L1, and 468-K18 in the Sanger centre database, for which PCR primers were designed. For each pair of primers, the 5' primer was labelled with one of the fluorescent dyes HEX, TET and FAM.

2.2.13.1 Preparation of samples

After amplification PCR products were checked on a 1% agarose gel, and then analysed on an ABI 377 machine. Several different microsatellites could be electrophoresed in the same lane provided they were labelled with different fluorescent dyes or were not of overlapping size ranges. Samples were prepared by diluting the PCR products, usually 1:10 with H_2O, and using 1.5µl of this dilution per lane, together with 0.5µl TAMRA 350 marker, and 1.5µl dextran blue/formamide dye. Before loading onto the gel, the samples were denatured at 95° for 5 minutes.

2.2.13.2 Running ABI gels

The gel was made using 25ml SeqaGel mix (National Diagnostics, 4.25%), and 200µl 10% ammonium persulphate was added shortly before pouring. The gel was allowed to polymerise for 2 hours before running. The wells were rinsed out thoroughly prior to running, to remove urea. Gels were run for 2 hours at 51°C using the ABI genescan run module 2400, with a constant voltage of 3kV, 40mW laser power, and 2400 scans per hour.

2.2.13.3 Analysis of results

The data was collected and analysed with the ABI Genescan software to determine the peak sizes for each sample.

2.2.14 Bisulphite modification and sequencing

The protocol for bisulphite modification of DNA detailed below is an adaptation of that used in the laboratory of Y. Hayashizaki, as reported in Kamiya *et al.* 2000 (Kamiya *et al.* 2000). Their method is itself a modification of the method used by (Clark *et al.* 1994).

One microgram of DNA was digested overnight at 37°C with 25 units of the restriction enzyme *Pvu*II in a total volume of 50µl. After digestion, the volume was increased to 100µl with distilled water, and then 10µl 3M sodium acetate, 1µl yeast tRNA, and 220µl 100% ethanol were added to precipitate the products of the digestion. The mixture was incubated on ice for 30 minutes, then centrifuged for 10 minutes at 14000 rpm in a microfuge. The pellet was washed with 70% ethanol, and then air-dried for 5 minutes, before resuspension in 50µl TE buffer (pH 8.0).

Denaturation was performed by adding 5.5µl 3M NaOH, and incubating at 37°C for 15 minutes, after which the bisulphite reaction was set up, by adding 30µl 10mM hydroquinone, and 520µl sodium bisulphite pH 5.0. The tubes were inverted a few times to mix, and then the reaction was overlaid with oil, and incubated in a 55°C water bath for 16 hours.

The products of the bisulphite reaction were purified with the geneclean kit from Bio101 utilising sodium iodide and glassmilk, and eluted into 20µl distilled water. A further denaturation was performed, by adding 2.2µl 3M NaOH, and incubating at 37°C for 15 minutes. The denatured bisulphite-modified products were precipitated, by the same method as described above, and then finally resuspended in 15µl TE buffer, pH 8.0. For PCR, 1µl was used per 50µl PCR reaction.

2.3 RNA METHODS

All RNA work was performed with full awareness of the possible risks of contamination with RNases. Appropriate precautions were taken to avoid these risks, by carrying out RNA manipulations away from the normal working area, using clean gloves for each manipulation, using a set of Gilson pipettes specifically designated for RNA work, and treating solutions with DEPC where possible. Autoclaved tubes and tips were used throughout. Samples of RNA were stored in a -70°C freezer, and only defrosted for sufficient time as to remove the amount required, before being quickly refrozen.

2.3.1 RNA extraction

RNA was extracted from the lymphoma samples using TRIzol reagent from GibcoBRL, and a protocol derived from the original method of (Chomczynski, Sacchi 1987).

The lymphoma tissue was rapidly minced on ice using a scalpel, then placed in a microcentrifuge tube containing 1ml of TRIzol reagent. The tubes were vortexed vigorously, several times over a 20 minute period, to break up the tissue as much as possible. To each tube was added 200µl chloroform, followed by 15 seconds of vigorous mixing. The tubes were incubated at room temperature for 2-3 minutes, and then centrifuged at 12,000×g for 15 minutes at 4°C. The upper aqueous phase containing the RNA was transferred to a fresh tube, and the RNA precipitated by adding 500µl isopropanol, and mixing well. Following a 10-minute incubation at room temperature, the tubes were centrifuged at 12,000×g for 10 minutes at 4°C. The supernatant was removed carefully and discarded. RNA washing was performed by adding 1ml of 75% ethanol, mixing well, and centrifuging at 7,500×g for 5 minutes at 4°C. The pellet was dried at room temperature for approximately 10 minutes, and then resuspended in 50µl DEPC-treated distilled water.

2.3.2 RNA quantification

RNA was quantified using a Genequant spectrophotometer.

2.3.3 RT-PCR

2.3.3.1 First strand cDNA synthesis

On ice, the following components were added to a nuclease-free tube: 10µl H_2O, 5µl primer (Race2), 3µl (1µg) RNA. The mixture was overlaid with mineral oil, incubated at 85°C for 3 minutes, then immediately cooled on ice for 2 minutes. To each tube was added 6µl 5× RTII buffer, 3µl 0.1M DTT, 1.5µl dNTP (stock containing 20mM each), 0.5µl (14.8 units) RNAguard, and 1µl (200units) Superscript™ II RT, then the tubes incubated at 42°C for 60 minutes, 50°C for 15 minutes, and 95°C for 15 minutes.

2.3.3.2 PCR

For PCR, 2µl of the RT reaction was used per 50µl PCR. Typically, a hot start was found to be necessary, and the success of the PCR was often improved by the addition of components such as 10% DMSO or 0.1%-1% Triton X-100.

2.3.4 Ribonuclease Protection Assay (RPA)

A ribonuclease protection assay was designed for the purpose of assessing quantitatively the amount of *ZAC* transcript in the lymphoma RNA samples. The transcription and ribonuclease protection assays were performed using the appropriate kits from Ambion.

2.3.4.1 Probe design

The RT-PCR product generated from a normal fetal cDNA by primers 1029-RT5'/3' was cloned into the vector pGEM®-T Easy, by methods as described in 2.2.10. Six of the resulting colonies were selected and miniprep cultures were grown, from which DNA was extracted by the SNAP™ miniprep kit method. Each clone was sequenced with the M13 reverse primer, to identify a clone in the correct orientation such that transcription from the T7 polymerase would generate an antisense transcript.

2.3.4.2 Transcription

The template for transcription was a PCR product generated from the clone mentioned above. To ensure that the template was RNase-free it was treated with 0.1% SDS and 200µg/ml proteinase K for 2 hours at 50°C, and then purified using the Wizard PCR purification kit. Of this, 100ng (1µl) was added to the transcription reaction, which also comprised 8.8µl nuclease-free H_2O, 2µl 10×buffer, 1µl each of 10mM ATP, CTP and GTP, and 4µM (3.2µl) 25µM [α^{32}P]-UTP. The transcription reaction was allowed to proceed for 15 minutes at 37°C, followed by DNaseI treatment, then removal of free nucleotides by ammonium acetate precipitation.

2.3.4.3 Calculation of yield and specific activity of probe

Yield and specific activity of the probe were calculated according to the percentage incorporation of [α^{32}P]-UTP as determined by TCA precipitation. The hybridisation reaction requires 2fmol of probe, so the probe was resuspended in a volume containing 2fmol per µl.

2.3.4.4 Hybridisation

Hybridisation of the probe and RNA samples was performed essentially as described in the manufacturer's recommended protocol. Reactions were set up containing

RNA and probe, which were then co-precipitated, resuspended, denatured, and then incubated overnight at 42°C. RNase digestion was performed at 37°C for 30 minutes, followed by precipitation. Samples were resuspended in gel loading buffer and electrophoresed on 5%, 8M urea denaturing polyacrylamide gels for 2 hours at 300 volts. Gels were exposed to autoradiographic film with intensifying screens at -70°C for 1-8 hours.

2.A APPENDIX

2.A.1 PCR primers

NAME	SEQUENCE 5'-3'
GENERAL	
M13Forward	gttttcccagtcacgacgttgta
M13 Reverse	agcggataacaatttcacacagga
T7	cgctaatacgactcactataggg
SP6	gtcgacatttaggtgacactata
Race2	gagctcgagtcgacatcga(t_{17})
CHAPTER 3	
p73-3	caggcccacttgcctgcc
p73-4	ctgtccccaagctgatga
1781-poly5'	gggccctgaagatccccgagcag
P73-RT3'	aagtgcacggcctccatgacccg
P73-RT5'	cccaagggttacagaccattta
79-T7-5'	caattgctctttattccaata
79-T7-3'(E2T7)	acagaaggctcgcactatcgt
79-SP6-5'	cacggtgaaaccccatctcta
79-SP6-3'	ggctcactcaactccaccttc
71-t7-5'	aggggttgatggagtgtcttgat
71-t7-3'	ctggccttccctcattgtctttt
71-sp6-5'	gtctggggctcctcacaactcta
71-sp6-3'	atacttacaaccgctctccaggg
286-t7-5'	catagcccctctctccagtg

286-t7-3'	cagcagaactgagaccacca
286-sp6-5'	gcctccagcattttccagta
286-sp6-3'	ctgcaccaggtaggtggaat
CPAN1	gggagctggaagcggagacag
CPAN2	ctggcccaaggtgagcagcac
CPAN3	tctgtggggcctgctcatcac
CPAN4	tctgaaatcgggactccaagc
CPAN5	gagccgaggacccgcaggaat
CPAN6	gcagctgtccatgtcaaaggg
CPAN0F	tccagaagcccaagagcgtga
CPAN3F	gtgatgagcaggccccacaga
CPAN5AR	cggtgtgcagaggcggct
CPAN 242-5'	gctggtgctgctcacctt
CPAN 242-3'	ctcgtgaaatgcactgagga
CPAN 321-5'	catttcacgagccacaggt
CPAN 321-3'	atgttctggctgacgttgtg
CPAN 678-5'	gacagaggagccaagggc
CPAN 678-3'	ggggttgatggagtgtcttg
CPAN 0AF	aggaagttcggcgtggctgg
CPAN 6F	ccctttgacatggacagctgc
CPAN 7R	attgcttccaccagtgtagg
CPAN 8R	cgggtcacagttgagcttgt
CPAN 9R	tgcttccgcttcaaccttgt
CPAN 4AF	ttcagagcaagtctggctatctg
CPAN 5BR	catggaccggagcttctggca
CPAN 1AR	ttgcgcagcacctcctggca
CPAN 9F	acaaggttgaagcggaagca
CPAN 677-int-R	gtccgcccgacacaaggtgtcc
CPAN 242-int-F	tgtggtcagaggctcttctt
CPAN 430-int-R	ttctggcacctcagactcca
CPAN 4F	cttggagtcccgatttcaga
cpan-ex1-pcr5'	ccagcttgcagagctcac
cpan-ex1-pcr3'	gctgaggcgaacgaaaacta
cpan-ex1-seq5'	atctgagcagctgggcagca
cpan-ex1-seq3'	cctattctccccacacgcct
cpan-ex2-pcr5'	tgtaaaacgacggccagtcagcctgagcctgcttcttta
cpan-ex2-pcr3'	caggaaacagctatgacctgagacccgagagttcacag
cpan-ex2-seq5'	agcacagctcattccggtcgt
cpan-ex2-seq3'	tgatgggcacctggagctaag
cpan-ex3-pcr5'	ctcaagtctgagtcctggtgatt
cpan-ex3-pcr3'	atgagcacattttcttccaagtc
cpan-ex3-seq5'	aggatgtgtcttcagctggaccg
cpan-ex3-seq3'	ttcttctggcacctcagactcca
cpan-ex4-pcr5'	gaggacagagcaagaccctg
cpan-ex4-pcr3'	agcacggagtgtggtcctac

cpan-ex4-seq5'	aggacacagacccagaccc
cpan-ex4-seq3'	acagagcctggcttcaaaaa
cpan-ex5-pcr5'	tgtaaaacgacggccagttggagtgagatggatcgaga
cpan-ex5-pcr3'	caggaaacagctatgaccagggcaagggctgaaggt
cpan-ex5-seq5'	gcgctgtgcaccaggctcaccga
cpan-ex5-seq3'	gctgtgttccacaggccagc
cpan-ex6-pcr5'	ggaggttgtagtaagccgagat
cpan-ex6-pcr3'	gaaacccaaggaggcaacac
cpan-ex6-seq5'	actccagcctggcgacagagcga
cpan-ex6-seq3'	acagggccctgcaggcactcgt
cpan-ex7-pcr5'	tgtaaaacgacggccagtactgtgactgcaatacactgc
cpan-ex7-pcr3'	caggaaacagctatgaccttaaaatgatgcccacgtca
cpan-ex7-seq5'	tgcctgtggcactgtcaccacag
cpan-ex7-seq3'	atgatgcccacgtcacgcctcaa
CPANRTPCR-5'	atcctcttcagcacctggaa
CPANRTPCR-3'	gtttccgcacaggctgct
CHAPTER 4	
NV149-STS2-5'	gcctatctggtatgaggtccac
NV149-STS2-3'	ggatgcggagactaacagaagt
NV149-seq-NotI-3'	acctcataccagataggcgcga
5108-PCR5'	gcagatcgtcacccgcaacc
5108-PCR3'	ctgccattacagtcaaccttc
5108-seqT3	ctagatcgggacagctgataatg
5108-seqT7	ctatgacctgagagacactgatg
5108seqF	gaaacacccagcttaggtcct
5108seqR	ccagatccttaactgttaaa
STSG9563-1	ctaaaaacataagacaatggag
STSG9563-2	gtaattcatcagtcattattagc
STSG24967-1	tggagaactctattgctccaca
STSG24967-2	ggctgtgtatttcttttcaggg
hzacpoly1-5'	ctgatgaaagagagcttgcaga
hzac2237-3'	gcaagtggggagtatgaggtag
LOT(A)-5'	ccccaggccactagataaaag
LOT(A)-3'	gaacagcttgagcaaatcaaga
ZAC-C-F1	ccatagcagatttgcctttgta
ZAC(D)-3'	gtgacagggaacatctgctg
ZAC(E)-5'	caggacatgccccagctat
ZAC(F)-3'	aaaggctttgccacagtca
hzac2018-5'	ctttccctttaggagcttctgc
hzacpoly1-3'	atagtcccagtctcgttttcca
PlagL1-5'	tggtgctgcaagtactgacc
PlagL1-3'	ggcagcaggcttatgcttta
ZAC(C)-5'	gcattttgggacaactgttttt
ZAC89-seq	tacaaaggcaaatctgctatgg

ZAC-RT5'	ggaatgttttcctagcttcattcccgtac
ZAC-621-3'	cttgcctatgatgtgtttccaa
ZAC-RT3' (2)	ggacctctcagctgtcactagct
PlagL1ex2r2	ctgaccaaatgctgtgccat
PlagL1ex2f1	tgattctgaagcggtcaggg
ZAC1029-RT5'	cagcagatgttccctgtcac
ZAC1029-RT3'	gaacttctccagggtgagga
CHAPTER 5	
D6S279-5'	Tet-gcttcaagtgtagggcaagaac
D6S279-3'	tggacctgggattatagggtag
D6S977-5'	Hex-gtgctcagcagcaggtgta
D6S977-3'	ctgtctttgtcaccaccagaaa
D6S1703-5'	Fam-ctggtgctgatgtatccaaaat
D6S1703-3'	tttggaggatcaggaaagaaaa
D6S308-F	Tet-gagagaattcacgtacataaacaca
D6S308-R	gtctaatcactgccccaaag
D6S311-F	Fam-atgtcctcattggtgttgtg
D6S311-R	gattcagagcccaggaagat
CA197-5'	Hex-tttatatgttgcatttcctttacaa
CA197-3'	caaaccatggcacacatatacc
CA340-5'	Tet-agagtttgcagtgagccaagat
CA340-3'	tctcctcactccctttcacttc
GT468-5'	Fam-tttttctttgtctggcttgtca
GT468-3'	agagaaaggcacaagctcaaaa
Bssu2	ggggtagtygtgtttatagtttagta
BSS12	craacacccaaacacctacccta
Bssu2-seqF	tagtygtgtttatagtttagt
SNRPN-M1	taaataagtacgtttgcgcggtc
SNRPN-M2	aaccttacccgctccatcgcg
SNRPN-P1	gtaggttggtgtgtatgtttaggt
SNRPN-P2	acatcaaacatctccaacaacca
GAPDH-F	cccatggcaaattccatggc
GAPDH-R	ccagccttctccatggtggt

2.A.2 Solutions

ABI gel loading buffer	95% deionised formamide, 20mM ethylenediamine tetra acetic acid (EDTA. disodium salt), 10mg/ml dextran blue

Acrylamide	30% Stock (19: 1) 28.5% acrylamide, 1.5% *N,N'*-methylenebisacrylamide
Alkaline Lysis Miniprep – Solution I	50mM glucose, 25mM Tris-HCl (pH 8.0), 10mM EDTA
Alkaline Lysis Miniprep – Solution II	1% SDS, 0.2M NaOH
Alkaline Lysis Miniprep – Solution III	3M potassium acetate, 2M acetic acid
Ammonium persulphate	10% or 25% w/v, in dH_2O
Ampicillin	50 mg/ml in H_2O stock solution, 20 µg/ml working concentration.
Denhardt's solution	5g Ficoll, 5g polyvinylpyrrolidone, 5g BSA, H_2O to 500ml
DEPC-treated water	Diethylpyrocarbonate added to dH_2O to a concentration of 0.1%, incubated overnight at 37°C, then autoclaved.
Digestion buffer for DNA extraction from human tissue	50mM Tris-HCl (pH 8.0), 100mM NaCl, 5mM EDTA, 1% SDS
Dimethylsulfoxide (DMSO)	100% stock
Ethidium bromide	10mg/ml in dH_2O Store in darkness at room temperature
First strand RTII buffer 5×	250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM $MgCl_2$
Gel loading buffer	0.25% bromophenol blue, 0.25% xylene cyanol FF, 15% Ficoll in water
Glycerol Tolerant Buffer	10× Stock 1.8M Tris, 1.7M taurine, 10mM EDTA
Hybridisation mix	5× Denhardt's, 5× SSC, 0.1% SDS, 0.1% sodium pyrophosphate, 20% dextran sulphate
Isopropylthio-β-D-galactoside (IPTG)	100mM in dH_2O

	Filter sterilize to 0.22 microns, store at -20°C
Kanamycin	10 mg/ml in H_2O stock solution, 10 µg/ml working concentration.
LB agar	15g Bacto-agar, 10g Bacto-tryptone, 5g Bacto-yeast extract, 5g NaCl, pH to 7.0 with NaOH, to 1 litre with dH_2O
LB/amp/X-Gal/IPTG plates	LB plates made with LB agar as above, supplemented with 100µg/ml ampicillin, 0.5mM IPTG and 80µg/ml X-Gal.
LB broth	10g Bacto-tryptone, 5g Bacto-yeast extract, 5g NaCl, pH to 7.0 with NaOH, to 1 litre with dH_2O
PCR buffer	10× buffer: 15mM $MgCl_2$, 100mM Tris-HCl (pH 8.0), 500mM KCl, in dH_2O
Phenol:chloroform:isoamyl alcohol	25:24:1 by volume
PNK buffer 10×	50mM Tris-HCl (pH 8.0), 10mM $MgCl_2$, 5mM DTT
Polyacrylamide gel – denaturing (5%)	28.8g urea, 6ml 10×TBE, 10ml 30% acrylamide stock (19:1), dH_2O to 60ml. Heat to dissolve. Polymerised with 300µl 10% ammonium persulphate and 64µl TEMED.
Polyacrylamide gel – non-denaturing (8%)	6.67ml 30% acrylamide (19:1), 2.5ml 10×TBE, dH_2O to 25ml, polymerised with 50µl 25% APS and 45µl TEMED.
Prehybridisation mix	5× Denhardt's, 5× SSC, 0.1% SDS, 0.1% sodium pyrophosphate
Quick Oligo Hybridisation Mix	0.5g bovine serum albumin (BSA), 0.5g

	polyvinylpyrrolidone , 0.5g Ficoll (type 400), 1g SDS, 1g sodium pyrophosphate, 250ml 20×SSC, 744ml dH_2O
Sequencing acrylamide 6% stock	200ml 30% Acrylogel 5, 19:1 stock, 480.5g urea, made up to 950ml with H_2O
Sequencing gels	95ml of 6% acrylamide stock and 5ml 20× glycerol tolerant buffer, polymerised with 100µl TEMED and 150µl 25% APS
Sodium dodecyl sulphate (SDS)	10% w/v, in dH_2O
Solution/Buffer	Components/Recipe
Standard saline citrate (SSC)	20× stock 3M NaCl, 0.3M trisodium citrate, pH to 7.0 with HCl
TE Buffer, pH 8.0	10mM TrisCl (pH 8.0), 1mM EDTA (pH 8.0)
TEMED (N,N,N',N'-tetramethylenediamine)	100% stock
Tris acetate electrophoresis buffer (TAE)	50×stock, per litre: 242g Tris base, 57.1ml glacial acetic acid, 100ml 0.5M EDTA (pH 8.0)
Tris borate electrophoresis buffer (TBE)	5×stock, per litre: 54g Tris base, 27.5g boric acid, 20ml 0.5M EDTA (pH 8.0)
X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside)	20mg/ml in dimethylformamide Store in darkness at -20°C

CHAPTER 3

CHROMOSOME 1p36, *TP73*, NEUROBLASTOMA, AND CANDIDATE TUMOUR SUPPRESSOR GENES

3.1 INTRODUCTION

The chromosomal region 1p36 has been implicated in a range of human tumours, but perhaps most widely studied to date is the significance of this region in neuroblastoma, a cancer of the neural crest accounting for 8% of all childhood malignancies. Genetic studies, including both classical cytogenetics and more recent techniques such as FISH, microsatellite analysis of loss of heterozygosity (LOH), and comparative genomic hybridisation (CGH) (Van Gele *et al.* 1997), have demonstrated that the two most common abnormalities specifically associated with neuroblastoma are 1p deletions, occurring in 30 to 40% of cases (Brodeur *et al.* 1981), and amplification of the oncogene *MYCN*, occurring in 25 to 30% of cases (Seeger *et al.* 1985; Brodeur *et al.* 1984).

MYCN is located at 2p24.1, and has similarity to the classic *MYC* oncogene (Schwab *et al.* 1983). Since its identification, amplification of *MYCN* has been identified in a large proportion of neuroblastomas, usually detectable by the presence of extrachromosomal double minutes, or homogeneously staining regions. A correlation between the amplification of *MYCN* and allelic loss of the short arm of chromosome 1 has been found in neuroblastomas, particularly those of a more aggressive nature (Schwab 1990; Fong *et al.* 1989). Indeed, *MYCN* has become a relatively reliable indicator of prognosis in neuroblastoma, with amplified cases being significantly associated with advanced stages of the disease (Brodeur *et al.* 1984).

3.1.1 Neuroblastoma suppressor genes at 1p

The frequent allelic loss that has been detected at 1p in neuroblastoma suggests that this region harbours a tumour suppressor gene whose loss is associated with the development of the disease. Mapping of the extent of the 1p deletions in neuroblastoma patients, and taking into account the correlation of only some 1p deletions with *MYCN* amplification has now revealed that there may be more than

one suppressor gene at 1p involved in neuroblastoma. Takeda *et al.* (Takeda *et al.* 1994), characterised 108 neuroblastoma tumours for loss of heterozygosity (LOH) at 1p. Of the 21 tumours that they identified to be informative for at least one marker, two distinct subgroups emerged from the analysis: firstly, those that had relatively small, interstitial deletions specifically at 1p36, had a good prognosis, and were not associated with *MYCN* amplification; and secondly, those that had much larger deletions extending from 1pter to 1p32, were much more aggressive tumours with a poorer prognosis, and had amplification of *MYCN* in the majority of cases. These findings led Takeda *et al.* to postulate that there could be two tumour suppressor genes at 1p, which are associated with clinically distinct classes of neuroblastoma. Similarly, Schleiermacher *et al.* (Schleiermacher *et al.* 1994) analysed LOH in 60 neuroblastoma patients, and detected two distinct consensus regions of deletion, of which only the more proximal one was associated with *MYCN* amplification. Furthermore, in agreement with these two studies, Cheng *et al.* (Cheng *et al.* 1995) characterised the deletions in seven cell lines without *MYCN* amplification and found them to be small, with the shortest region of overlap at 1p36.23-33, whereas in nine cell lines with *MYCN* amplification there were larger deletions extending from 1p35-36.1 to the telomere. As established in the studies by Takeda *et al* and Schleiermacher *et al.*, the overall conclusions from this study are that there are two suppressor genes at 1p, of which only one is associated with *MYCN* amplification.

3.1.2 One of the putative neuroblastoma suppressor genes is likely to be imprinted

In a study by Caron *et al.* (Caron *et al.* 1993), a significant observation was made with regard to the parental origin of the lost allele in 1p deleted cases of neuroblastoma: in 13 of 15 patients with LOH at 1p36, the lost allele was of maternal origin. This study thus provided the first evidence of imprinting at 1p. However, the results of a study by Cheng *et al.* (Cheng *et al.* 1993) did not support this idea, since they found the allelic loss at 1p in neuroblastoma to be of random parental origin. In 1995, Caron *et al.* (Caron *et al.* 1995) studied the extent of deletions and parental origin of allelic loss in neuroblastomas with and without *MYCN* amplification. They

reported a significant difference in the results between the two groups, such that *MYCN* single-copy tumours had a small consensus region of deletion at 1p36.2-3, and the allelic loss was of preferential maternal origin, whereas the *MYCN* amplified cases showed the characteristic large deletion as observed previously, but the allelic loss was of random parental origin. This evidence indicates again that there are two tumour suppressor genes at 1p, but for the first time suggests that the more distal gene at 1p36 is imprinted, whereas the proximal gene is not. Contrary to the study of Cheng *et al.* (Cheng *et al.* 1993), Caron *et al.* (Caron *et al.* 1995) distinguished between *MYCN* amplified and non-amplified cases of neuroblastoma, and in doing this they were able to identify the imprinting effect of only one of the putative suppressors, whereas this was masked in Cheng's study. Thus, by analysing the parental origin of the allelic loss separately in *MYCN* amplified and non-amplified cases, the conflicting evidence from previous studies was reconciled.

It is now apparent from these results that this chromosomal region harbours two distinct tumour suppressor genes. The first, which is more proximal and is eliminated by large deletions extending from the telomere to 1p35, may have some kind of role as a repressor of *MYCN* amplification. Its deletion would therefore eradicate this function, allowing amplification of the *MYCN* oncogene, with tumorigenic consequences. Neuroblastoma patients with *MYCN* amplification and deletions of this putative suppressor typically have tumours of a clinically aggressive nature, and have poor prognoses. The second putative suppressor gene is located more distally at 1p36.23-33, and is likely to be a maternally expressed imprinted gene, since there is preferential maternal loss at this locus. This tumour suppressor is not associated with the regulation of *MYCN* amplification, and patients with deletions confined to this region have much better survival prospects.

3.1.3 Candidates for the imprinted suppressor gene

Focussing in on the putative imprinted tumour suppressor locus, the identification of candidate genes in this region is clearly an important goal. However, the criteria to be met in identifying this tumour suppressor are stringent: the evidence has indicated

that this suppressor is a maternally expressed imprinted gene, and it must therefore be shown that it is the maternal allele that is lost in deletion cases. However, since not all neuroblastoma cases are associated with 1p deletions, it must also be expected that in at least some non-deletion cases the candidate gene function is eliminated by other means, such as mutation of the expressed allele. Additionally, the neuroblastoma suppressor would be expected to reside in the consensus region of deletion at 1p36.23-33, and possess an appropriate function which, when eliminated, could feasibly be responsible for tumour development. To justify further investigation, candidate genes must be known to satisfy at least some of these criteria.

To date, several genes have been proposed as candidates for the putative imprinted tumour suppressor at 1p36, mostly on the basis of their genomic location. These candidates include *ID3*, an inhibitor of DNA binding found to exhibit very low levels of expression in neuroblastoma (Ellmeier *et al.* 1992); *TNFR2*, a tumour necrosis factor receptor gene (White *et al.* 1993); *CDC2L1* (*PITSLRE*), a cell division control-related protein kinase gene deleted in some neuroblastoma cell lines (Lahti *et al.* 1994); and *DAN*, (differential screening-selected gene aberrant in neuroblastoma), a transcription factor (Enomoto *et al.* 1994). Attempts to define the extent and precise location of the region of deletion in neuroblastoma had narrowed down this region to a 72cM or 28Mb interval, but could not delineate it any more accurately (Fong *et al.* 1989; Weith *et al.* 1989; Takayama *et al.* 1992; Takeda *et al.* 1994; Schleiermacher *et al.* 1994). However, a study of LOH in a large panel of primary neuroblastoma tumours, and analysis of deletions and breakpoints in neuroblastoma cell lines enabled White *et al.* (White *et al.* 1995) to refine this interval to 39cM or 8Mb. The identification of this consensus region believed to contain the putative suppressor gene has eliminated the four proposed candidates mentioned above, on the basis that they lie outside the critical boundaries of the interval, suggesting that the true gene remained to be identified. A good remaining candidate is the gene human kruppel-related 3 (*HKR3*), a zinc-finger transcription factor gene which maps to 1p36.3, and is expressed at particularly high levels in fetal and adult nervous tissue (Maris *et al.*

1997). However, the lack of mutations detected in this gene in neuroblastoma tumour DNAs and cell lines argues against a tumour suppressor role.

3.1.4 *TP73*, a candidate neuroblastoma suppressor gene

Perhaps the best proposed candidate to date is *TP73*, which has been shown to be maternally monoallelically expressed in some tissues, and whose protein product appears to have cell cycle control functions analogous to those of the structurally similar *p53* protein (Kaghad *et al.* 1997). Indeed, the protein encoded by *TP73* shares significant amino acid similarity with the three functional domains of the p53 protein, with 29%, 63% and 38% identity between their transactivation, DNA binding and oligomerization domains, respectively (Kaghad *et al.* 1997). Overproduction of p73 results in transcriptional activation of p53-responsive genes such as *CDKN1A* (also known as $p21^{waf}$), and causes inhibition of cell growth by induction of apoptosis, in a manner reminiscent of the action of p53 (Jost *et al.* 1997).

Since its initial characterisation in 1997, *TP73* has been pursued as the putative 1p36 tumour suppressor gene, not only in neuroblastomas, but also in a wide range of other tumours associated with this region. Does *TP73* fulfil the criteria? The crucial factors that led to its proposal as a candidate for the neuroblastoma tumour suppressor gene were:

1. The location of *TP73* at 1p36.33 – a region which is frequently deleted in a wide range of cancers, but particularly neuroblastoma.
2. High amino acid similarity of the protein product of *TP73* to that of the known tumour suppressor gene *TP53,* and the ability to activate p53-responsive genes in order to induce apoptosis.
3. Monoallelic expression of *TP73* in neuroblastoma and non-neuroblastoma cell lines, and normal blood.

Considering the combination of these factors, it seemed logical to hypothesise that *TP73* was the putative neuroblastoma suppressor gene at 1p36. Being imprinted effectively gives *TP73* its first 'Knudson hit', since only one allele is expressed. To achieve a complete loss of function therefore demands that only the single expressed allele is lost by mutation or deletion. By losing the sole expressed allele of *TP73*, the function of the protein in inducing apoptosis would be effectively eliminated. A loss of control such as this could be expected to result in tumorigenesis.

Following the initial characterisation of the *TP73* gene, subsequent research has focussed on analysing the imprinted status of *TP73*, and attempting to identify tumour-specific mutations that might account for the expected 'second hit'.

Contrary to expectation, mutation screening efforts in a diverse range of tumour types associated with allelic loss at 1p36 have so far identified very few mutations in *TP73*. It is apparent that there are several common polymorphisms in the genomic sequence of *TP73*, as these have been detected in many of the mutation screening studies described below, but very few tumour-specific alterations have been identified that could be envisaged to have a specific role in tumorigenesis.

Neuroblastoma has been an obvious focus for mutation screening analyses. Kovalev *et al.* (Kovalev *et al.* 1998) screened the entire open reading frame in 16 neuroblastoma tumours but detected only common polymorphisms, and no tumour-specific alterations. Ichimiya *et al.* (Ichimiya *et al.* 1999) analysed 140 primary neuroblastomas by SSCP, and found only two tumours with amino acid substitutions in the C terminal domain that were likely to result in a loss of function. These substitutions were a proline to arginine change, and a proline to leucine change, both at residues within the domain of the protein that is expected to have a role in transactivation. Ejeskar *et al.* (Ejeskar *et al.* 1999), and Han *et al.* (Han *et al.* 1999) analysed 30 and 23 neuroblastomas respectively, but detected only common polymorphisms in the gene.

In lung cancer, Mai *et al.* (Mai *et al.* 1998a) and Nomoto *et al.* (Nomoto *et al.* 1998) analysed 21 and 61 lung cancer cases respectively, detecting only silent nucleotide substitutions and polymorphisms that were present in both normal and tumour samples. Yoshikawa *et al.* (Yoshikawa *et al.* 1999) identified substitutions and deletions that affected the amino acid sequence in 3 of 17 lung cancer cell lines. One of these cell lines had a glycine to tryptophan substitution. The alterations in the other two cell lines were a 12bp deletion in exon 13, resulting in a substitution of one amino acid and deletion of the next four; and deletions of 2bp and 4bp in exon 10, resulting in the deletion of two amino acids but retaining the reading frame. Yoshikawa *et al.* speculated that these could be cancer related alterations, despite not having analysed any normal samples. In hepatocellular carcinoma, Mihara *et al.* (Mihara *et al.* 1999) were unable to detect any tumour specific mutations in 48 untreated tumour samples; Peng *et al.* (Peng *et al.* 2000) however did identify one apparently tumour-specific 5-nucleotide deletion when they analysed 22 pairs of tumour and matched normal samples. Mutational analysis of *TP73* was performed in prostatic cancer by Takahashi *et al.* (Takahashi *et al.* 1998), and Yokomizo *et al.* (Yokomizo *et al.* 1999a), but the few genetic alterations detected were also found to be present in normal samples, and were considered to be neutral sequence polymorphisms. In colorectal cancer Sunahara *et al.* (Sunahara *et al.* 1998) and Han *et al.* (Han *et al.* 1999), used SSCP techniques to analyse 82 and 43 tumour samples respectively, and as with other studies identified common polymorphisms present in both tumour and normal samples, but failed to identify any tumour-specific alterations in the sequence. Similarly, in gastric cancer, Kang *et al.* (Kang *et al.* 2000), and Han *et al.* (Han *et al.* 1999) identified only polymorphisms in their mutation screening efforts. To analyse bladder tumour samples for the presence of *TP73* mutations, Yokomizo *et al.* (Yokomizo *et al.* 1999b) employed the denaturing high-performance liquid chromatography technique (DHPLC), but identified only polymorphisms that had been previously documented. Bladder cancer was also scrutinised for mutations in *TP73* by Chi *et al.* (Chi *et al.* 1999), but their SSCP analyses did not reveal any mutations. Breast cancer, for which a high frequency of allelic loss at 1p has been established, was screened for mutations by Shishikura *et al.* (Shishikura *et al.* 1999), Zaika *et al.* (Zaika *et al.* 1999), Schwartz *et al.* (Schwartz

et al. 1999), and Han *et al.* (Han *et al.* 1999), but apart from several common polymorphisms, revealed only one genetic alteration that resulted in an amino acid substitution.

The range of cancer types studied for the presence of mutations in the *TP73* gene also includes melanoma (Kroiss *et al.* 1998; Schittek *et al.* 1999; Tsao *et al.* 1999), ovarian cancer (Ng *et al.* 2000), oesophageal cancer (Nimura *et al.* 1998), oligodendrogliomas (Mai *et al.* 1998b; Tsujimoto *et al.* 2000), and Merkel cell carcinomas (Van Gele *et al.* 2000), but again very few tumour-specific genetic alterations have been identified. Whilst the tumour suppressor gene *TP53* is mutated frequently in most of these cancers, *TP73* is rarely subject to alterations that could cause a loss or alteration of its function. From the extensive scale of the mutation screening studies described above, and the diverse range of tumour types that have been analysed, it is justifiable to conclude that mutations of this gene are very infrequent. Thus, the possibility that mutation could act as a second 'hit' in the inactivation of *TP73* in a classic tumour suppressor model is unlikely.

With regard to the imprinting of *TP73*, despite the initial report of monoallelic expression by Kaghad *et al.* (Kaghad *et al.* 1997), a significant proportion of the evidence generated by subsequent studies has been contradictory to this, revealing biallelic expression of *TP73* in both fetal and adult normal tissues (Hu *et al.* 2000; Yokomizo *et al.* 1999a; Yokomizo *et al.* 1999b; Kovalev *et al.* 1998; Chen *et al.* 2000; Peng *et al.* 2000). With few exceptions, the results of expression analysis generally fall into one of two categories:

1. The *TP73* gene is imprinted, and shows monoallelic expression in normal tissues. However, rather than finding loss of the expressed allele in cancer, which would be expected in a two-hit suppressor gene model, it was shown that *TP73* was biallelically expressed in tumours. In effect, the transition from monoallelic to biallelic expression of *TP73* was associated with the development of cancer. In renal cell carcinoma, Mai *et al.* (Mai *et al.* 1998c) investigated the expression status of *TP73* using a polymorphism in exon 2,

and demonstrated monoallelic expression in eleven of twelve normal tissues, but biallelic expression in eight of the twelve corresponding tumour samples. They confirmed monoallelic expression using a different polymorphism in two informative families, and were able to attribute the expression to the maternal allele in fetal pancreas and thymus. Thus, their results are in agreement with the original reports of maternal-specific monoallelic expression (Kaghad *et al.* 1997), and suggest that the loss of imprinting and resultant biallelic expression of the gene is a common event that may be associated with renal cell carcinoma. Using the same polymorphism, Chi *et al.* (Chi *et al.* 1999) found that 52.2% of heterozygous bladder cancer specimens expressed *TP73* biallelically, whereas four non-cancerous samples did not, although these were not matched pairs. They did however find tumour-specific biallelic expression in one matched pair, suggesting that the activation of a silent allele of the gene may be related to the progression of cancer. In gastric adenocarcinoma, Kang *et al.* (Kang *et al.* 2000) demonstrated monoallelic expression in 19 non-cancerous samples, and biallelic expression of 23.8% of the corresponding cancer samples, suggesting that the biallelic expression of *TP73* may be involved in a subset, but not all cases of gastric adenocarcinoma. Similarly, a switch from monoallelic expression to biallelic expression was identified in 50% of neuroblastomas (Liu *et al.* 2000b), and 27% of oesophageal squamous cell carcinomas (Cai *et al.* 2000), suggesting that this mechanism may be relevant in some, but not all cancers.

Conversely:

2. The *TP73* gene is not imprinted, and shows biallelic expression in both normal and tumour DNA. However, the level of expression of *TP73* is found to be significantly increased in tumours as compared to matched normal samples. Studies that fall into this category include those by (Kovalev *et al.* 1998) (Yokomizo *et al.* 1999b) and (Chen *et al.* 2000b). Kovalev *et al.* found biallelic expression in three of three normal tissues and five of six

neuroblastomas, but identified elevated expression of the *TP73* gene, such that tumours overexpressed the gene up to 90-fold more than normal tissues. Yokomizo *et al.* found biallelic expression of *TP73* in both normal and bladder cancer samples, thus refuting evidence for imprinting of this gene, yet they detected an average 6-fold overexpression in 22 of the 23 samples analysed, with as much as a 20-fold increase in some samples. Chen *et al.* identified biallelic expression in 70.8% of normal samples and 91.7% of ovarian cancer samples, suggesting that although a minority may undergo a switch from monoallelic expression to biallelic expression, the majority express the gene biallelically in both normal and cancerous states. However, overexpression of the gene was correlated to advanced ovarian cancer rather than earlier stages or borderline ovarian tumour, indicating that the elevated expression of the gene may be a factor in the progression of tumours.

Clearly, whether *TP73* is actually imprinted remains questionable, perhaps due to a complex pattern of tissue-specific, and time-specific imprinting, and even variability between individuals. However, what is of more interest from these findings, is that irrespective of which of the above scenarios holds true, the consequences are very similar – it would appear that overexpression of *TP73* is associated with tumorigenesis, whether by transition from monoallelic to biallelic expression in the tumour, or merely by upregulation of the normal expression levels of the biallelically expressed gene. Indeed, most of the studies described above in which a transition from monoallelic to biallelic expression was detected, also demonstrated increased expression of the gene, which in these cases is likely to be attributable to the switch from the expression of one allele to the expression of two. As suggested by Zaika *et al.* (Zaika *et al.* 1999), since the imprinting of *TP73* is so "idiosyncratic", and a consistent pattern has not been established, it is of more relevance to look for tumour-specific changes in expression levels of *TP73*. Along these lines, there are several studies that have not investigated the imprinting status of *TP73*, but instead have assessed only the expression levels and found a higher level of *TP73* expression in tumours as compared to normal DNA controls, for example Sunahara *et al.* found increased *TP73* levels in colorectal carcinomas as compared to normal controls

(Sunahara *et al.* 1998); Tannapfel *et al.* detected elevated *TP73* levels in a subset of hepatocellular carcinomas (Tannapfel *et al.* 1999), and Tokuchi *et al.* found that in 87% of lung cancers *TP73* was expressed at more than twice the level of the matched normal sample (Tokuchi *et al.* 1999).

Unexpectedly therefore, and in contradiction with the originally proposed hypothesis of *TP73* acting as a tumour suppressor gene, it is now apparent that the involvement of *TP73* in cancer is related to its upregulation, rather than its loss. This finding is substantiated by the recent report that transgenic mice deficient in *TP73* do not have an increased susceptibility to spontaneous tumorigenesis (Yang *et al.* 2000), indicating therefore that *TP73* is not acting as a tumour suppressor gene. If, conversely, it is the upregulation of *TP73* that is associated with cancer, what is the mechanism by which this overexpression causes cancer? One proposed mechanism is that the p73 protein plays a role in controlling p53 function, by inhibiting its DNA binding (Vikhanskaya *et al.* 2000). The experiments that resulted in this postulate involved transfection of wild type p73 into a p53-expressing ovarian cancer cell line, which had the effect of significantly decreasing the transcriptional activity of p53. Similar experiments transfecting the same cell line with a mutant version of p73 (containing a defect in the DNA binding site) did not cause the decrease in p53 transcriptional activity. The conclusions from this study were that p73 modulates p53 function by inhibiting its DNA binding, and in this way, overexpression of *TP73* could be a novel mechanism of inactivation of *TP53*.

In accepting that *TP73* is highly unlikely to be the putative neuroblastoma tumour suppressor gene at 1p36, future research must be directed at both understanding the alternative role of *TP73* in cancer, but perhaps more importantly, continuing the search for the tumour suppressor gene which remains undiscovered in this chromosomal region. To use *TP73* as a starting point for the identification of further candidate genes is justifiable for two reasons: firstly, *TP73* resides in the consensus region of deletion as determined by deletion mapping studies in neuroblastoma, and secondly, there is evidence for imprinting in this region, which, although conflicting, cannot be disregarded. Since imprinted genes tend to be clustered together in the

genome, it is probable that other imprinted genes will be found in this region. The purpose of this study therefore was to systematically investigate the region surrounding *TP73*, to determine whether other genes might reside close to *TP73*, and whether such genes might be imprinted and could be considered as candidates for the putative neuroblastoma suppressor gene.

3.2 RESULTS

3.2.1 Isolation and initial characterisation of PAC clone 79-L8, containing the *TP73* gene

To begin to analyse the region surrounding *TP73*, it was necessary to obtain a PAC clone containing the gene, since only the cDNA sequence of *TP73* was available at this time. Using primers p73-3 and p73-4, as described in (Kaghad *et al.* 1997), a product of 482bp representing part of exon 2 of *TP73* was generated from genomic DNA. Sequencing confirmed that it was the correct product. This PCR product was labelled and used as a hybridisation probe for identifying a PAC clone from the high-density gridded filters of the RPCI1 human PAC library from HGMP. A single clone, 79-L8 was identified.

In an attempt to identify putative CpG islands within the PAC clone 79-L8, DNA was prepared from the clone, and analysed by digestion with *Not*I, *Asc*I, and double digestion with *Not*I and *Asc*I together. The recognition sequences of the *Not*I and *Asc*I restriction enzymes are GCGGCCGC and GGCGCGCC respectively. These are rare-cutting enzymes, and any sites for these enzymes are highly likely to reside within a CpG island due to the presence of the CpG dinucleotides within the recognition sequences.

The products of digestion were separated on a pulsed field gel, shown in Figure 4. It appears that there are five fragments generated by digestion with *Not*I, of approximate sizes 55kb, 45kb, 18kb, 16kb and 15kb. The 16kb fragment corresponds to the expected size of the vector fragment, since the insert is cloned into a *Not*I site. To distinguish between the similar sized fragments of 15kb and 16kb, an unrelated PAC clone was also digested with *Not*I, and run alongside PAC clone 79-L8; the band that is common to both clones represents the vector, thus confirming that the other fragment of 15kb is part of the insert of PAC 79-L8. The fragments resulting from digestion with *Asc*I are of approximate sizes 65kb and 85kb. These

results indicate that there are five *Not*I sites in 79-L8 (of which three are internal, and two are the end sites), and two *Asc*I sites. Additionally, the results indicate that the total insert size of this PAC clone is approximately 150kb.

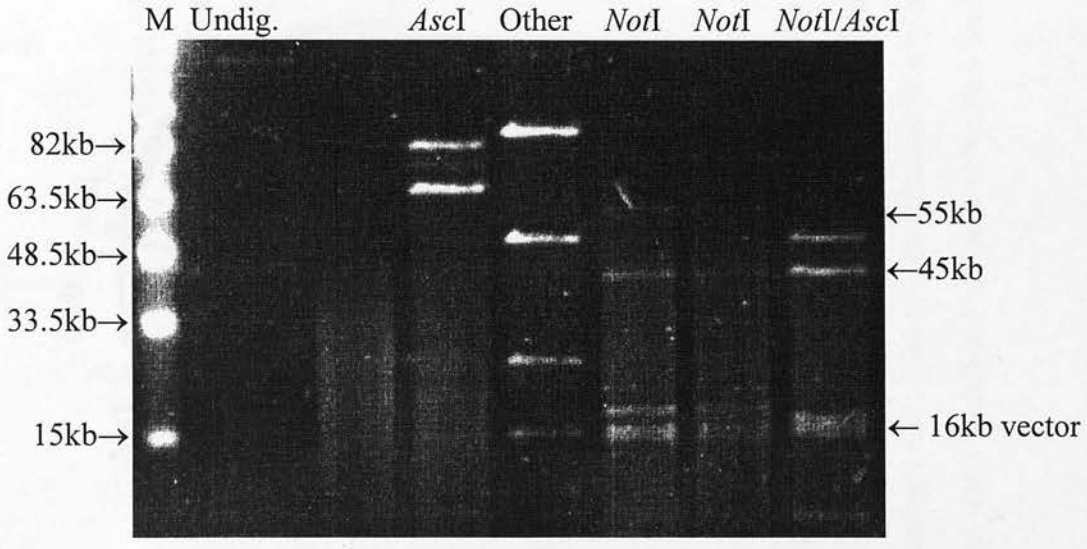


Figure 4. 1% PFGE gel of restriction digests of PAC clone 79-L8. M is the NEB midrange PFG marker, with bands of sizes indicated on the left. Undig. is a sample of undigested PAC clone DNA, to determine the clone size. The PAC clone was digested singly with *Asc*I, *Not*I, and doubly with both *Not*I and *Asc*I. The lane marked 'Other' is a *Not*I digest of an unrelated PAC clone, to assist in identification of the common 16kb vector band. Electrophoresis conditions were 6V/cm, 14°C, with switch times between the alternating electric fields ramped from 1 to 10 seconds during a 24 hour period.

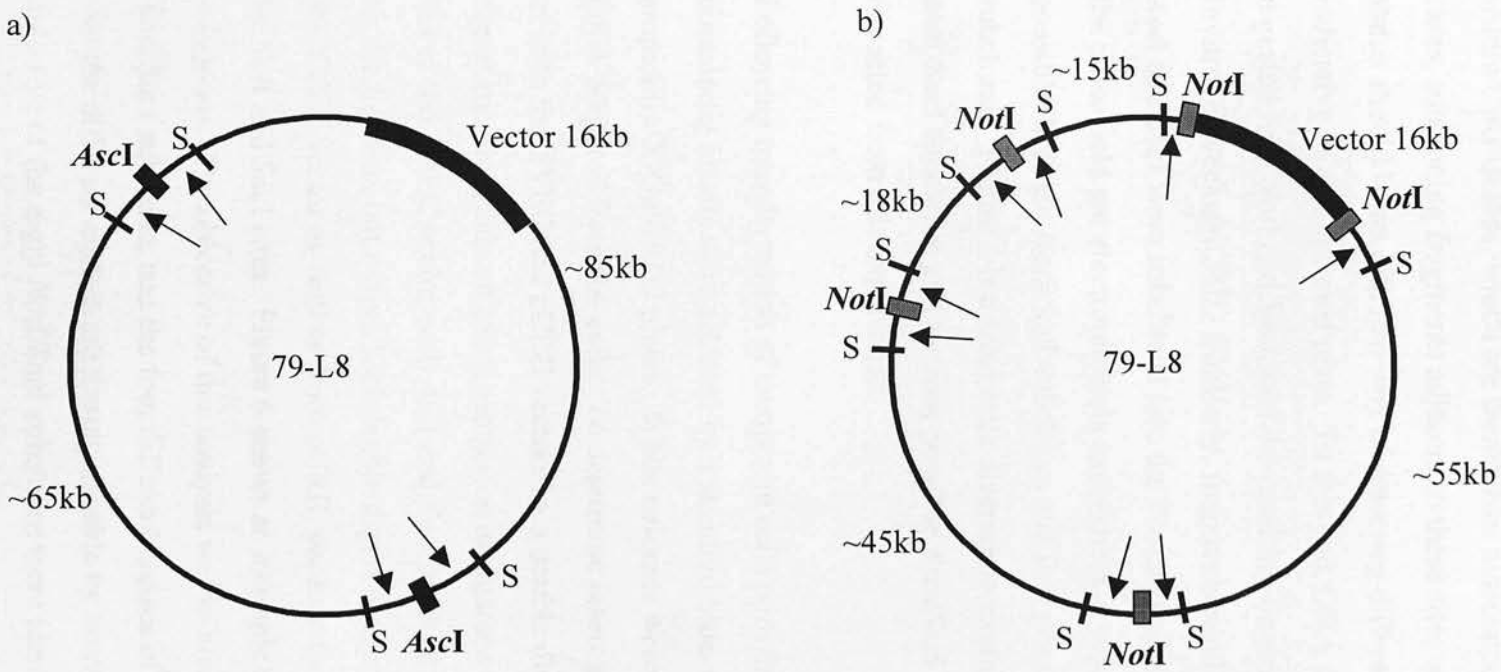


Figure 5. Diagram to explain the possible subclones that would be generated by a) *Asc*I-*Sac*I subcloning, and b) *Not*I-*Sac*I subcloning of PAC clone 79-L8. The approximate sizes between *Asc*I sites (a), or *Not*I sites (b) are given, based on the PFGE data. It was not possible at this stage to deduce the actual order of the fragments, so the *Asc*I and *Not*I sites are presented in an arbitrary order. S denotes a putative *Sac*I site. The arrows indicate the expected subclones in each case. There are 4 possible *Asc*I-*Sac*I subclones, and 8 possible *Not*I-*Sac*I subclones (including the two end clones which represent the *Not*I site in the vector).

3.2.2 Subcloning the *Not*I and *Asc*I sites of PAC clone 79-L8 for methylation status analysis

Since the sites recognised by the restriction enzymes *Not*I and *Asc*I almost always lie within CpG islands, which are themselves associated with genes in at least 60% of cases, subcloning fragments adjacent to these sites and analysing their methylation status should be an efficient way of detecting differential methylation, which may be indicative of an imprinted gene. To this end, DNA from PAC 79-L8 was double-digested with *Not*I and *Sac*I, and the resulting fragments subcloned into the Invitrogen vector pCRII. Similarly, fragments resulting from double-digestion with *Asc*I and *Sac*I were subcloned into the Promega vector TVEC. From the results of the pulse field gel electrophoresis experiment, it would be expected to obtain eight possible different *Not*I/*Sac*I subclones and four possible different *Asc*I/*Sac*I subclones. Figure 5 is a schematic diagram to explain the origin of the eight possible *Not*I/*Sac*I subclones and the four possible *Asc*I/*Sac*I subclones that would be expected from this experiment.

Following transformation of competent cells with the ligated vector, colonies containing inserts were selected by a standard blue-white selection procedure using ampicillin/X-Gal/IPTG plates. White colonies were picked, grown overnight, and DNA prepared from the cells. To determine insert size, the inserts were digested out of both the TVEC and pCRII vectors by a double digest with *Not*I and *Sac*I, and the digest run alongside an uncut sample on an agarose gel. Inserts cannot be digested out of the TVEC vector with *Asc*I and *Sac*I because the *Mlu*I cloning site is destroyed by the ligation, but a *Not*I/*Sac*I double digest is appropriate to release the insert from the TVEC vector as well as from pCRII, because the cloning site of TVEC is flanked by *Not*I and *Sac*I sites. Figure 6 shows an example of this analysis for seven of the subclones. The objective of this analysis was to identify the eight different classes of *Not*I/*Sac*I subclone, and the four different classes of *Asc*I/*Sac*I subclone, on the basis that the different classes are distinguishable by insert size. After several attempts, only five of the eight *Not*I/*Sac*I subclones were identified, including the two *Not*I subclones representing the *Not*I sites at the ends of the clone, and three others

arbitrarily named P3, P20, and P25. All four of the *Asc*I/*Sac*I subclones, named T1, T3, T8 and T9, were identified. Sequencing of the beginning of the inserts of the subclones with the M13 forward primer enabled the sequences to be checked against the BLAST database. There were problems associated with the techniques in this experiment, in that approximately 25-30% of subclones were found to contain inserts derived from *E.coli* sequences. Additionally, it would seem that preferential subcloning of some of the *Not*I/*Sac*I fragments was occurring, to the extent that some fragments were never subcloned. This may have been due to a bias in insert size preference of the subcloning system under these conditions.

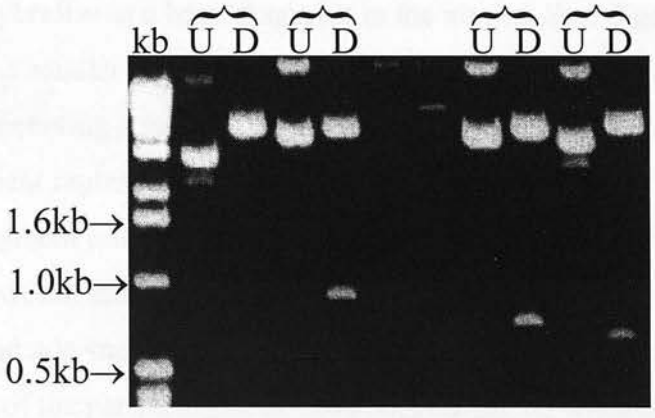
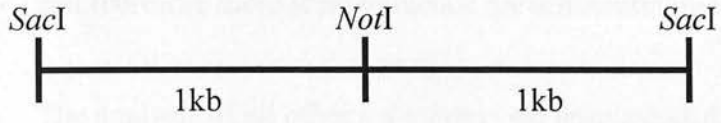


Figure 6. Analysis of insert size following subcloning of *Not*I-*Sac*I digest fragments of PAC clone 79-L8 into the vector pCRII. The gel shows the analysis of insert size for four subclones. Each pair of lanes represents a sample of the undigested (U) and digested (D) subclone DNA. Digests were performed using *Not*I and *Sac*I together. Sizes of the 1kb ladder bands are given on the left.

To analyse the methylation status of the subclones T1, T3, T8, T9, P3, P20 and P25, genomic Southern blots were prepared for each of these subclones, comprising both normal and parthenogenetic DNA digested singly with *Sac*I, and doubly with either *Not*I and *Sac*I, or *Asc*I and *Sac*I accordingly. The insert of each of the subclones was used as a probe to hybridise to the appropriate Southern blot. The purpose of this was to detect if differential methylation was present between the paternal and maternal alleles. For example, if the *Not*I site associated with a *Not*I/*Sac*I subclone was differentially methylated on the two parental alleles, the probe would be expected to hybridise to a large fragment in the normal *Sac*I digest, but to both this fragment and a smaller fragment in the normal *Not*I/*Sac*I digest (the large *Sac*I fragment representing a methylated allele which prevents *Not*I digestion, and the smaller fragment representing an unmethylated site, where *Not*I has digested the large *Sac*I fragment into two smaller fragments, one of which is detected by the probe). This would indicate differential methylation. The parental origin of the methylated and non-methylated alleles would be apparent from the equivalent hybridisation of the parthenogenetic blots, depending on whether the probe hybridised to the larger (methylated) or the smaller (unmethylated) fragment. Figure 7 is a theoretical diagram to explain the rationale behind this experiment.



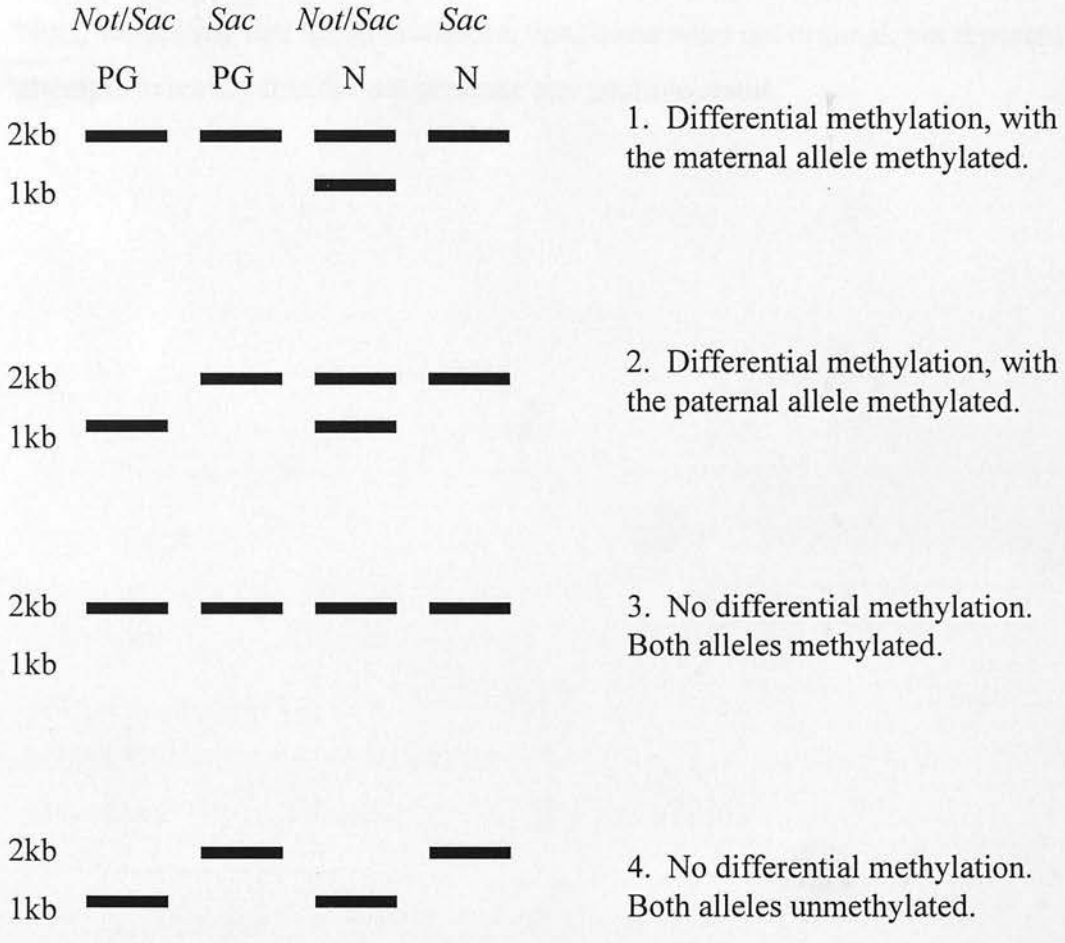


Figure 7. Diagram to explain the Southern blot analysis of methylation status. A hypothetical 2kb *Sac*I fragment containing a *Not*I site is shown at the top. The visual result of the four possible scenarios with respect to the methylation status of the *Not*I site is given in 1-4. PG = parthenogenetic DNA. N = normal genomic DNA.

The subclones T1, T3, T8, T9, P3, P20 and P25 were analysed by this method. The results of T1 and T3 are shown in Figure 8. In the case of T1, the *Asc*I site is completely methylated, and in the case of T3, the *Asc*I site is completely unmethylated. The pattern is the same for both normal and parthenogenetic DNA, and therefore there is no evidence for differential methylation.

The analysis of all other subclones was unsuccessful. This was due to the presence of repeats in the probe sequence in the case of T9 and P3, which resulted in non-specific binding of the probe to the filter thus obscuring the genuine result. In the case of subclones T8, P20 and P25, the probes did not appear to hybridise to the blots, suggesting that the hybridisation conditions were not optimal, yet repeated attempts to rectify this did not generate any positive result.

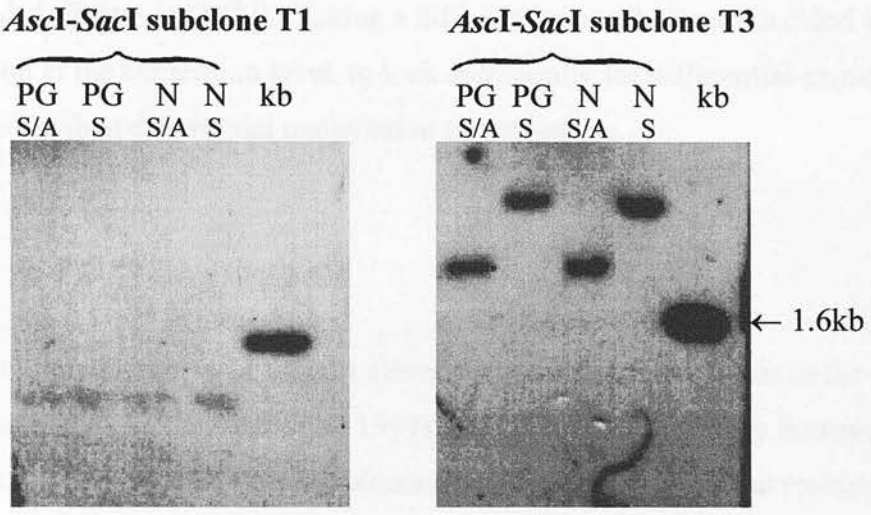


Figure 8. Analysis of methylation status of *Asc*I sites in PAC clone 79-L8. The inserts of subclones T1 and T3 were used to probe genomic Southern blots of digested parthenogenetic (PG) and normal (N) control DNA, digested with *Sac*I and *Asc*I together (S/A), or with *Sac*I alone (S). The *Asc*I site associated with subclone T1 is methylated, whereas the *Asc*I site associated with subclone T3 is unmethylated. There is no evidence of differential methylation between the maternal and paternal alleles in either case. The 1.6kb band of the kilobase ladder marker is visible.

The methylation analysis of the *Not*I and *Asc*I sites of PAC clone 79-L8 did not show any evidence of differential methylation, although as described above, the experiments were unsuccessful in some cases due to the presence of repeat sequences. Additionally, not all the sites were successfully subcloned, so the results are incomplete and do not represent a full analysis of the methylation status of all *Not*I and *Asc*I sites in 79-L8. Taking a different approach, it was decided to analyse the region at the expression level, to look specifically for differential expression of *TP73* rather than differential methylation of the region.

3.2.3 Is *TP73* imprinted?

There are now known to be at least eleven common polymorphisms in the coding sequence of *TP73* (Schwartz *et al*. 1999). At the time of this study however, there were only four identified polymorphisms in *TP73*. One of these, at residue 1781 of the cDNA sequence accession number Y11416 was chosen for RT-PCR experiments because it alters a site for the restriction enzyme *Bst*UI, such that the site is present when the polymorphic base is 'A', but is removed if the polymorphic base is 'G'. Taking advantage of this fact, it was possible to identify informative fetuses for this polymorphism to use in an RT-PCR experiment.

A panel of 66 first trimester fetal DNAs was available, along with corresponding maternal DNAs for the majority of them. Paternal DNAs were not available. Details of this panel are given in Table 1 in Materials and Methods. The rationale behind this RT-PCR experiment is that if a fetus can be identified that is heterozygous for the *TP73* 1781 polymorphism, with a mother who is homozygous for the same polymorphism, then the parental origin of both alleles in the fetus is known, since the non-maternal allele in the heterozygous fetus must be derived from the father. Performing RT-PCR on RNA from the heterozygous fetus enables the expression status of *TP73* to be analysed with respect to the parental origin of the expressed allele(s).

3.2.3.1 Polymorphism 1781

The panel of fetal DNAs was first genotyped with respect to the polymorphism at residue 1781, by using primers 1781POLY5' and p73-RT3', to generate a PCR product of size 179bp, followed by digestion of the PCR products with the restriction enzyme *Bst*UI, and electrophoresis of the resulting fragments on 3% NuSieve agarose gels. Figure 9 shows the location of these primers, in relation to the polymorphic *Bst*UI site. The PCR product contains the polymorphic *Bst*UI site, and an additional, non-polymorphic *Bst*UI site, which functions effectively as an internal control for complete digestion. When digested, the resulting fragments comprise a 67bp fragment in all cases, and a 112bp fragment which is intact when the polymorphic residue is G, but is further digested into two fragments of 85bp and 27bp when the polymorphic residue is A.

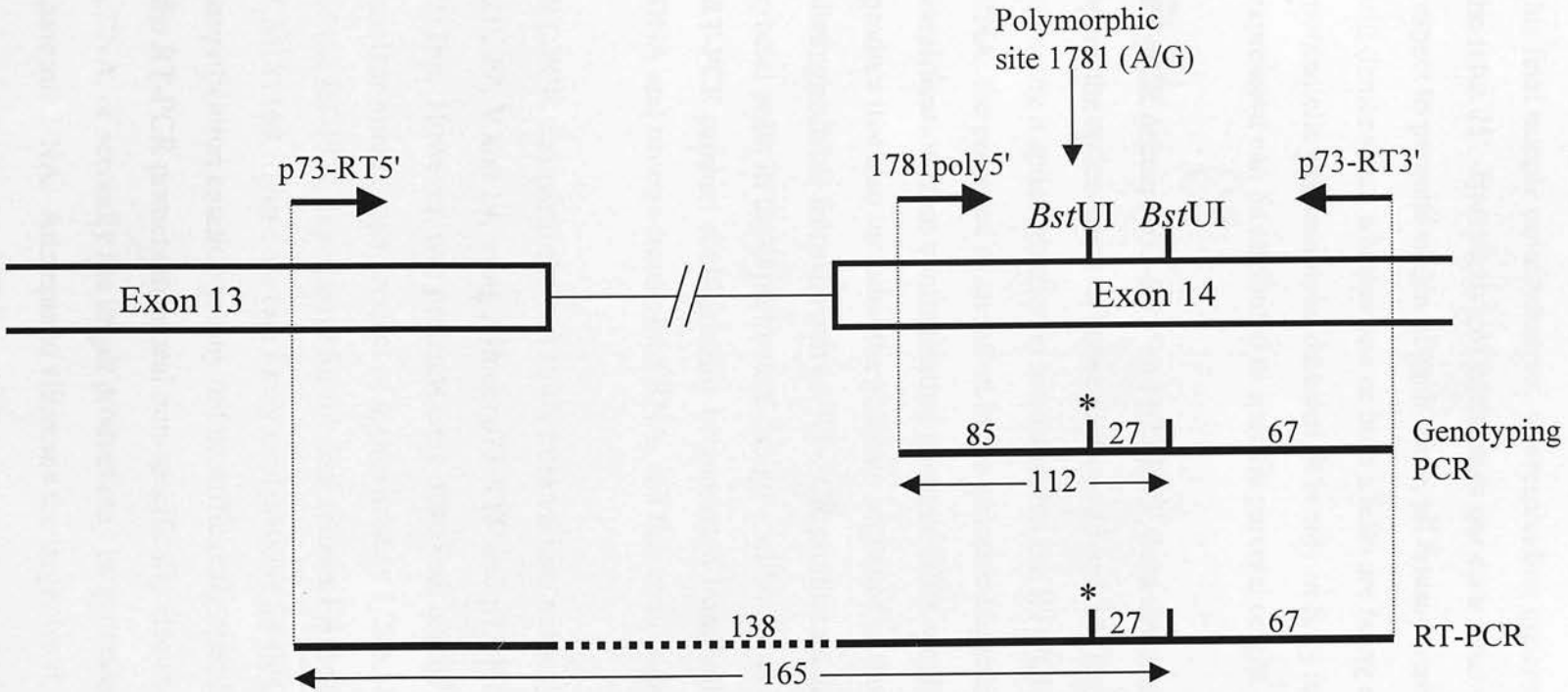


Figure 9. Diagram of the genotyping PCR and RT-PCR for the polymorphism at residue 1781 of *TP73*. Primers 1781poly5' and p73RT-3' generate a product of 179bp for the purpose of genotyping the fetal DNAs for the polymorphism. Primers p73RT-5' and p73RT-3' generate a 232bp RT-PCR product from cDNA but not from genomic DNA since the product spans a splice junction. There are two *Bst*UI sites; the polymorphic site which is present only when the polymorphic residue is an A rather than a G, is shown with an asterisk. The sizes of the resulting fragments following *Bst*UI digestion are shown in base pairs. Not to scale.

Four heterozygous (G/A) fetuses were identified, known as 21' (FB11), 29' (SG28), V (JF101272), and 14 (CM181169). Maternal genotyping by the same procedure revealed the corresponding maternal DNAs of 29' and 14 to be homozygous (G/G) for the 1781 polymorphism. Maternal DNA V was heterozygous, thereby rendering this fetal sample uninformative. Corresponding maternal DNA was unavailable for the fetus 21'. Irrespective of whether or not each fetus was fully informative with respect to parental origin of each allele, all fetuses were analysed, since RT-PCR will still demonstrate whether one or both alleles are being expressed. However, should monoallelic expression be detected, it is only in fully informative cases that expression can be attributed to specific parental origin.

RT-PCR primers p73-RT5' and p73-RT3' were designed so that the RT-PCR product spans the splice junction between exons 13 and 14 (Figure 9.) The reason for spanning a splice junction is to ensure that the RT-PCR product originates only from RNA; the presence of an intron in the genomic sequence means that any amplification from contaminating genomic DNA would generate a much larger product that also included the genomic sequence of the intron, which would be easily distinguishable from the correct RT-PCR product on the basis of size. This is a crucial point in the experimental design – without this feature, an identically sized RT-PCR product could feasibly be generated from both contaminating genomic DNA and reverse-transcribed RNA, and this could significantly distort the results.

RT-PCR was performed on RNA from various tissues of the heterozygous fetuses 21', 29', V and 14, using primers p73-RT5' and p73-RT3', to give a product of size 232bp. However, two products were observed; one of the expected size 232bp, and another much larger product of approximately 1.2kb. Figure 10 shows the products of the RT-PCR experiment for the four fetuses FB11, SG28, JF101272 and CM181169. There are two likely explanations for this result. Firstly, the amplification conditions may not be sufficiently specific, therefore allowing one of the RT-PCR primers to anneal non-specifically elsewhere on the reverse-transcribed cDNA, or secondly the larger product may be generated from contaminating genomic DNA. Attempts to eliminate the larger band, by both DNaseI pre-treatment

of the RNA, and PCR optimisation were not effective. However, the successful RT-PCR products from the heart, muscle, lung, eye, and kidney of fetus JF101272 were less affected by this problem (Figure 10), and were digested by *Bst*UI and electrophoresed on a 3% NuSieve agarose gel, shown in Figure 11. Digestion of the RT-PCR products with *Bst*UI should cleave the 232bp product into fragments of 67bp and 165bp if the polymorphic base is A, but the 165bp fragment is cleaved further into two fragments of 138bp and 27bp if the polymorphic base is G. In all five tissues of this fetus, the 232bp RT-PCR product was digested into fragments of 165bp, 138bp and 67bp (and presumably a band of 27bp which is not visible here), a result which suggests that both alleles are expressed, and that *TP73* is biallelically expressed in these fetal tissues, at least within this individual.

To confirm the presence of both alleles in the RT-PCR product, the 232bp band derived from heart (fetus JF101272) was excised from the gel, purified, and sequenced on both strands. At the polymorphic site, both G and A residues can be seen (Figure 12), thus confirming the result of the digests described above.

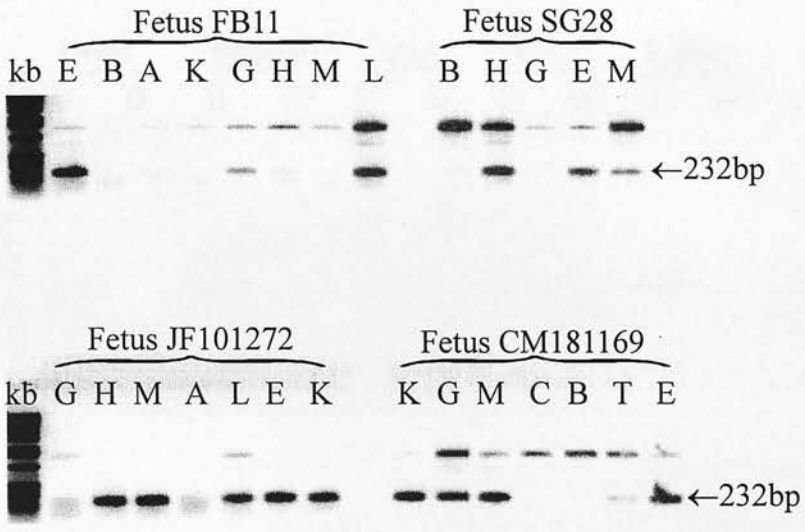


Figure 10. RT-PCR products for the *TP73* polymorphism at residue 1781, for four different fetuses. The tissues are as follows: E = eye, B = brain, A = adrenal, K = kidney, G = gut, H = heart, M = muscle, L = lung, C = cord, T = testes. The smaller product is of the expected size of 232bp.

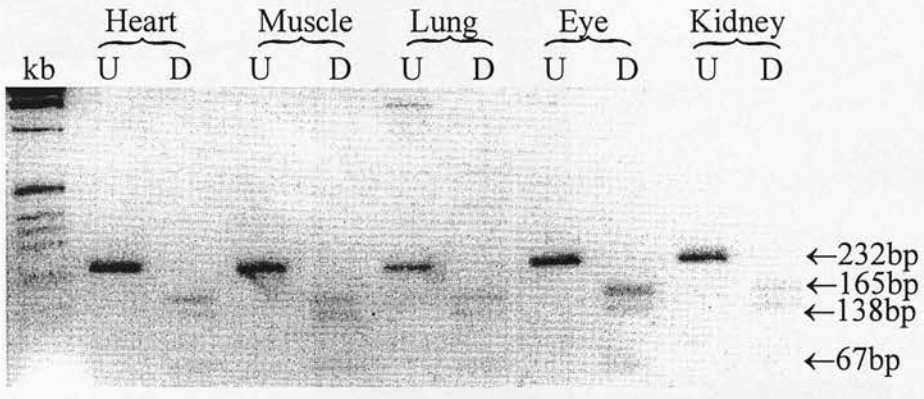


Figure 11. *TP73* RT-PCR products for heart, muscle, lung, eye and kidney, of fetus JF101272, digested with *Bst*UI and electrophoresed on a 3% agarose gel alongside corresponding undigested RT-PCR product. The uncut RT-PCR product is 232bp. Figure 9 explains the expected digestion products for heterozygous and homozygous individuals. Note that the 27bp fragment is not resolved on this gel. The sizes of the visible bands are given on the right.

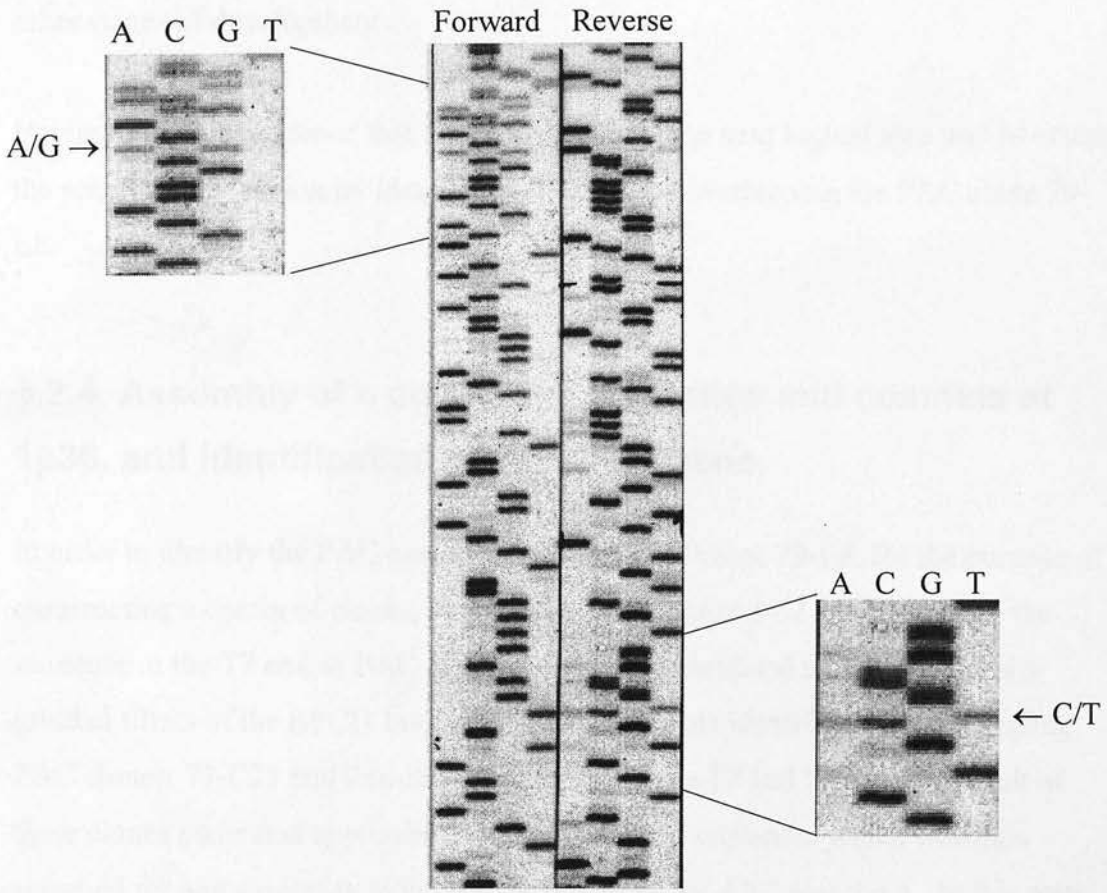


Figure 12. Example of *TP73* RT-PCR product derived from fetal heart, sequenced in both forward and reverse directions to demonstrate that both parental alleles are present at the polymorphic residue 1781 (A/G on the forward strand; C/T on the reverse strand), and that expression of *TP73* is therefore biallelic in this case.

In summary, this study did not reveal any evidence to suggest that *TP73* is imprinted in the heart, muscle, lung, eye or kidney of first trimester fetuses. Expression is apparently biallelic in these tissues according to the results presented here. However, this does not rule out the possibility that *TP73* may be imprinted in other tissues, or other stages of development.

Having found no evidence that *TP73* is imprinted, the next logical step was to extend the search of this region by identifying PAC clones overlapping the PAC clone 79-L8.

3.2.4 Assembly of a contig of PAC clones and cosmids at 1p36, and identification of the *DFFB* gene.

In order to identify the PAC clones adjacent to PAC clone 79-L8, for the purpose of constructing a contig of clones, an oligonucleotide named E2-T7, specific to the sequence at the T7 end of PAC clone 79-L8, was hybridised to the high density gridded filters of the RPCI1 human PAC library. This identified two overlapping PAC clones; 71-C21 and 286-D6. Sequencing at the T7 and SP6 ends of each of these clones generated approximately 200-300bp of sequence, which was then searched for any similarity to known sequences by BLAST searching. In this way, the T7 sequence of PAC clone 71-C21 was found to be apparently identical to part of a known published cDNA encoding the recently identified gene *CPAN/CAD/DFFB* (Halenbeck *et al.* 1998), (which will be known as *DFFB* in this study). The BLAST match is shown in Figure 13. The identity between the genomic sequence at the T7 end of PAC clone 71-C21, with the cDNA sequence of *DFFB*, spans a region of 93bp, after which the sequences diverge from each other. This point of divergence is likely to represent a splice junction in the cDNA sequence. This is shown in Figure 13.

Since the PAC clones 79-L8 and 71-C21 containing respectively the genes *TP73* and *DFFB* are overlapping, it is likely that these genes reside no more than 200kb apart.

```
gb|AF039210|AF039210 Homo sapiens caspase-activated nuclease mRNA, complete
cds
           Length = 1017

 Score =  184 bits (93), Expect = 8e-45
 Identities = 93/93 (100%)
 Strand = Plus / Minus


71C21-T7:1    TTCCAGGTGCTGAAGAGGATCCTGCTCTCCCTGTTACTGTAGGGGTTGATGGAGTGTCTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DFFB     :770 TTCCAGGTGCTGAAGAGGATCCTGCTCTCCCTGTTACTGTAGGGGTTGATGGAGTGTCTT 711


71C21-T7:61   GATAAGCAGCTGTCCATGTCAAAGGGACCctgggggccaaaagacaatgagggaaggcca..
              |||||||||||||||||||||||||||||
DFFB     :710 GATAAGCAGCTGTCCATGTCAAAGGGACCctgg--------------------------
```

Figure 13. BLAST output for a sequence generated at the T7 end of PAC clone 71-C21. Residues 1-93 of this sequence (reading from the end of the clone into the insert) are a perfect match for residues 678-770 of the published cDNA sequence of caspase-activated nuclease (*CPAN/DFFB*). The genomic sequence of the PAC clone diverges from the cDNA beyond residue 678, suggesting that there is a splice junction in this region. The exonic sequences are shown in upper case; the presumed intronic sequence is shown in lower case.

The gene *DFFB* encodes a 40kDa protein that induces DNA fragmentation after activation by caspase-3, in response to apoptotic stimuli. The protein was first described in the mouse, as DFF (DNA fragmentation factor), by Liu *et al.* (Liu *et al.* 1997). They showed that it exists in a heterodimeric complex of 40kDa and 45kDa subunits. The larger of the two subunits is cleaved at two sites by caspase-3 to generate an active factor, which has the capacity to induce DNA fragmentation and chromatin condensation during apoptosis (Liu *et al.* 1998). The same protein was identified and characterised in the mouse by another group, who called it caspase activated DNase (CAD). Similarly, they found that the complex comprised CAD together with an inhibitor protein ICAD that contains cleavage sites for caspase-3 (Enari *et al.* 1998; Sakahira *et al.* 1998). The human homologue of the DFF or CAD/ICAD complex was first described by Halenbeck *et al.* (Halenbeck *et al.* 1998), who referred to it as caspase activated nuclease (CPAN). Their experiments demonstrated that as a nuclease, CPAN is sufficient to degrade naked DNA and induce apoptosis in naïve nuclei. Mukae *et al.* (Mukae *et al.* 1998), cloned the human CAD independently, and showed it to be expressed in a variety of tissues including pancreas, spleen, prostate and ovary. They also concluded that CAD is responsible for DNA fragmentation during apoptosis, and that the gene encoding CAD/CPAN/DFF40 maps to human chromosome 1p36.3. To summarise, *DFFA* is the gene that encodes ICAD/DFF45, which was previously mapped to 1p36.3 (Leek *et al.* 1997); *DFFB* encodes CAD/CPAN/DFF40, and also maps to 1p36, close to *TP73*.

To assemble a contig of the PAC clones 71-C21, 79-L8, 286-D6, their relative orientations were determined. A short STS was designed at both the T7 and SP6 ends of each of the PAC clones 71-C21, 79-L8 and 286-D6, for the purpose of orientating each clone with respect to the others. Performing these STS PCRs on each of the PAC clones and also on a cosmid, 176-g8 or 44-N16 (identification of this cosmid is described in detail later) revealed which ends overlapped the other clones (Figure 14a), and a contig was assembled by making deductions from this information (Figure 14b).

The physical and chromosomal relationship between some of these clones was examined by FISH/fibre FISH (performed by Dr. F.Speleman and his group). Firstly, since *DFFA* has also been shown to map to the chromosomal region 1p36.2-36.3 (Leek *et al.* 1997), the positions of these two genes relative to the centromere were determined. This was performed using dual colour FISH, with the PAC clone 92-K14 (found to contain *DFFA* by Leek *et al.* (Leek *et al.* 1997)), and the cosmid clone 176-g8 which contains *DFFB*. These probes were distinguishable on prometaphase chromosomes, with *DFFB* mapping distal to *DFFA*, with respect to the centromere. An amino acid comparison of DFFA and DFFB reveals significant similarity, particularly at the N-terminus. This is shown in Figure 15. Considering the close proximity of these genes to each other, it is likely that they evolved at least partly through a gene duplication mechanism.

Secondly, fibre FISH was also used to determine the percentage overlap of clone 71-C21 with each of the other clones 79-L8, 286-D6 and cosmid 176-g8. The results of this were as follows, with the overlap of 71-C21 given as a percentage of the length of the other clone: 79-L8: 23%; 286-D6: 55%; 176-g8: 19%. These results confirm the clone orientations already established by STS analysis as shown in Figure 14, and provide additional quantitative data about the extent of each overlap. The results of the FISH analysis are illustrated in the paper attached in the Appendix at the end of this thesis (Judson *et al.* 2000).

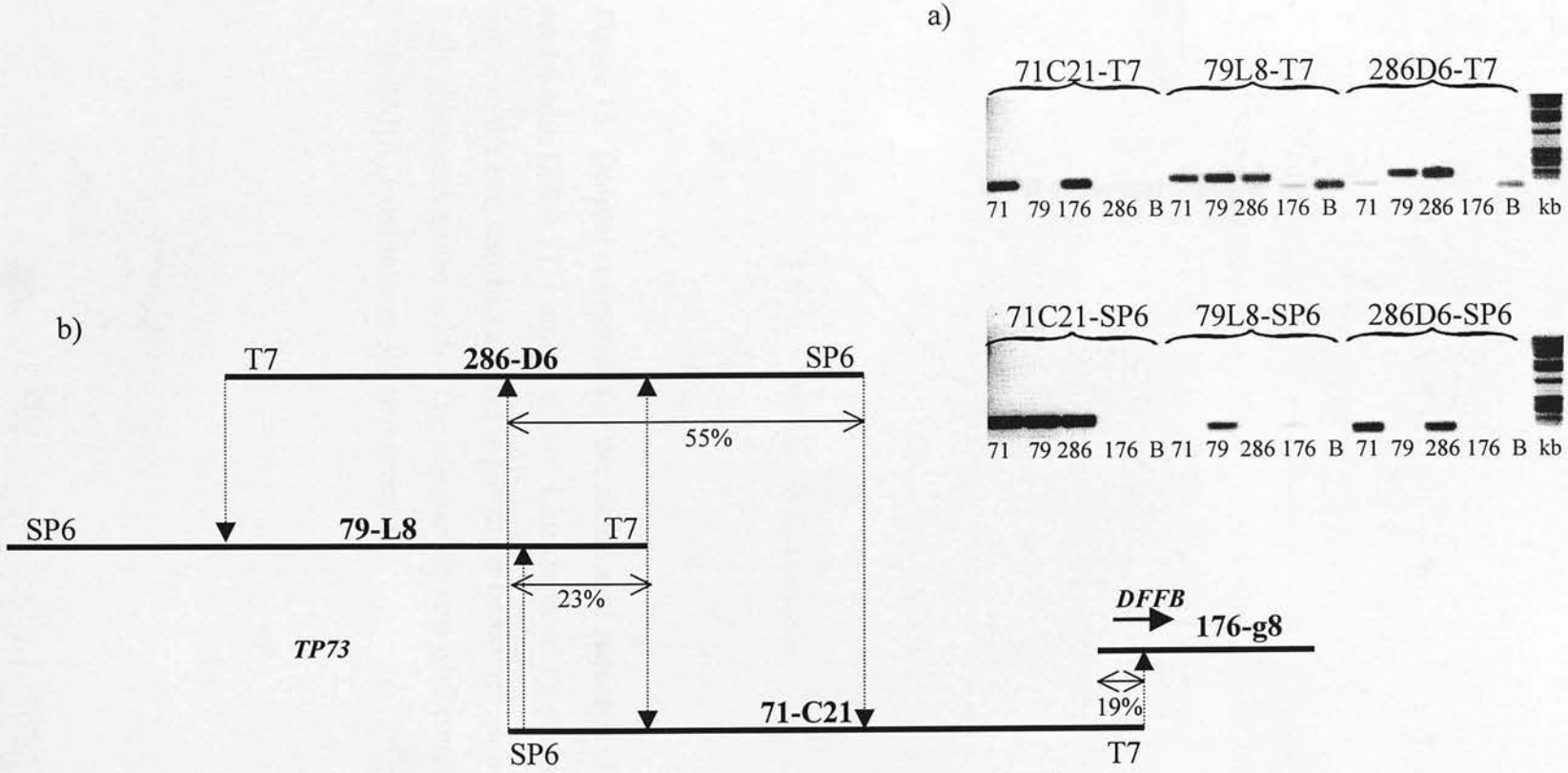


Figure 14. a) STS PCRs for the T7 and SP6 ends of PAC clones 71-C21, 79-L8, and 286-D6, performed on DNA from each of these PAC clones and cosmid 176-g8, to determine their orientation and overlap with respect to each other. b) The deduced orientation of the clones according to the data from these STS PCRs. The location of the *DFFB* gene is shown. The overlap of PAC clone 71-C21 with the other clones (from fibre FISH analysis; see Figure 2 in attached paper Judson *et al*. 2000) is shown as a percentage of their length.

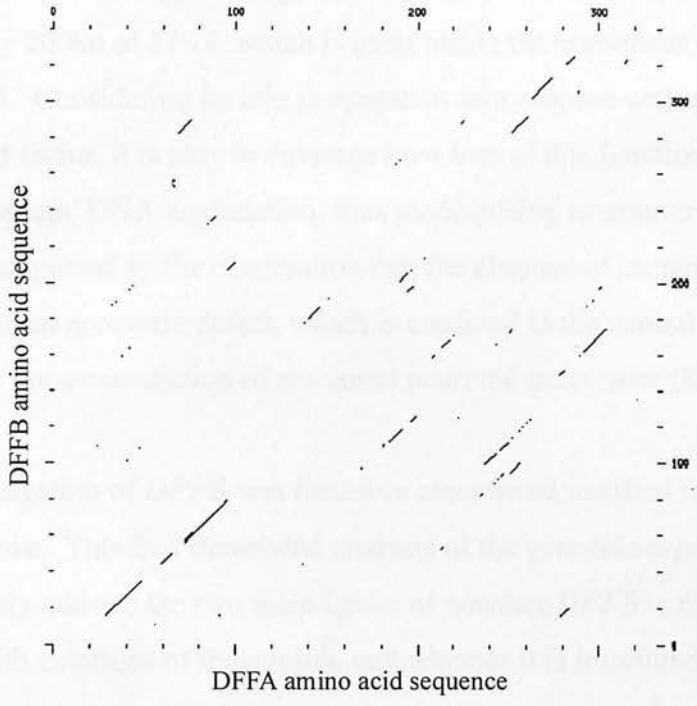


Figure 15. Dotplot comparison of the amino acid sequence of DFFB (339 amino acids) with DFFA (331 amino acids). Lines represent regions of significant identity between the two sequences; there is particular homology at the N terminus. The scale represents amino acids. The comparison was performed using UWGCG COMPARE (window size 30, stringency 11).

3.2.5 *DFFB* – a candidate tumour suppressor gene for neuroblastoma?

On the basis of its chromosomal location and putative biological role, *DFFB* already satisfies some of the criteria that must be fulfilled by candidates for the neuroblastoma tumour suppressor gene (see 3.1). *DFFB* is likely to reside within approximately 200kb of *TP73*, which is itself inside the consensus region of deletion at 1p36.23-33. Considering its role in apoptosis as a caspase-activated DNA fragmentation factor, it is easy to envisage how loss of this function would result in failure of apoptotic DNA degradation, thus predisposing to tumour development. This idea is supported by the observation that the absence of caspase-3 function in mice results in an apoptotic defect, which is confined to the central nervous system, and results in the accumulation of abnormal neuronal precursors (Kuida *et al.* 1996).

Further investigation of *DFFB* was therefore considered justified for the reasons described above. This first demanded analysis of the genomic organization, in order to subsequently address the two main issues of whether *DFFB* is mutated in tumours associated with deletions of this region, and whether it is imprinted.

3.2.6 Analysis of *DFFB* gene structure

No information had been published regarding the gene structure of *DFFB*, except for a partial cDNA sequence (coding region only), which is reproduced in Figure 16 (Halenbeck *et al.* 1998). The BLAST match of *DFFB* residues 770–678 with residues 1-93 of the T7 sequence of PAC clone 71-C21 suggests that there is a splice between residues 677 and 678, because the PAC sequence diverges from the cDNA at this point, and because the genomic sequence shows a consensus splice acceptor sequence (Figure 13). In order to locate other splice junctions, sequencing primers (CPAN1 – CPAN6) were designed approximately every 120bp in the cDNA, and used for sequencing directly on PAC clone 71-C21. By analysing where the resulting sequences diverged from the known cDNA sequence, splice junctions were

identified. For each new splice junction that was identified, flanking primers were designed, and used to determine the approximate intron size by PCR on both PAC 71-C21 DNA and genomic DNA. These primers are named CPAN 0F, 3F, 5AR, 242-5', 242-3', 321-5', 321-3', 678-5', 678-3', 0AF, 6F, 7R, 8R, 9R, 4AF, 5BR, 1AR, 9F, 677-int-R, 677-int-F, 430-int-R, and 4F, and their sequences are given in Appendix 2.A.1, Materials and Methods. Five splice junctions were identified, suggesting that *DFFB* has six coding exons.

```
  1 atgctccaga agcccaagag cgtgaagctg cgggccctgc gcagcccgag gaagttcggc
 61 gtggctggcc ggagctgcca ggaggtgctg cgcaagggct gtctccgctt ccagctccct
121 gagcgcggtt cccggctgtg cctgtacgag gatggcacgg agctgacgga agattacttc
181 cccagtgttc ccgacaacgc cgagctggtg ctgctcacct tgggccaggc ctggcagggc
241 tatgtgagcg acatcaggcg cttcctcagt gcatttcacg agccacaggt ggggctcatc
301 caggccgccc agcagctgct gtgtgatgag caggccccac agaggcagag gctgctggct
361 gacctcctgc acaacgtcag ccagaacatc gcggccgaga cccgggctga ggacccgccg
421 tggtttgaag gcttggagtc ccgatttcag agcaagtctg gctatctgag atacagctgt
481 gagagccgga tccggagtta cctgagggag gtgagctcct acccctccac ggtgggtgcg
541 gaggctcagg aggaattcct gcgggtcctc ggctccatgt gccagaagct ccggtccatg
601 cagtacaatg gcagctactt cgacagagga gccaagggcg gcagccgcct ctgcacaccg
661 gaaggctggt tctcctgcca gggtcccttt gacatggaca gctgcttatc aagacactcc
721 atcaacccct acagtaacag ggagagcagg atcctcttca gcacctggaa cctggatcac
781 ataatagaaa agaaacgcac catcattcct acactggtgg aagcaattaa ggaacaagat
841 ggaagagaag tggactggga gtatttttat ggcctgcttt ttacctcaga gaacctaaaa
901 ctagtgcaca ttgtctgcca taagaaaacc acccacaagc tcaactgtga cccgagcaga
961 atctacaaac cccagacaag gttgaagcgg aagcagcctg tgcggaaacg ccagtga
```

Figure 16. The cDNA sequence of DNA fragmentation factor 45/caspase-activated nuclease (*DFFB*), from Genbank Accession Number AF039210. The start and stop codons are shaded.

Two disadvantages of using PAC clone 71-C21 became apparent. Firstly, direct sequencing on this clone was not consistently successful, perhaps due to the quality or amount of the DNA. Secondly, the analysis could not be completed, because PAC 71-C21 contains only part of the *DFFB* gene. All of the sequence which is 3' of base 771 of the cDNA is absent from the PAC clone. To overcome both of these problems, a cosmid clone was obtained, which contained the entire *DFFB* gene. This was done by generating a 311bp genomic PCR product (with primers 6F and 9R) spanning residues 685 to 995 of the cDNA, for use as a hybridisation probe for the filters of the LL01NC01 chromosome 1 cosmid library obtained from HGMP-RC. This PCR product was chosen as the probe because it covers a substantial region of the cDNA, and is located in the part of the gene that is absent from PAC clone 71-C21. It would therefore be expected to identify a clone containing at least the missing region of the cDNA, if not the entire *DFFB* gene. A single cosmid clone was identified by the hybridisation; 176-g8 (also known as 44-N16.)

NB. As will be explained later, the 6F/9R PCR product used as probe was actually, with hindsight, likely to be derived from a chromosome 9 *DFFB* pseudogene. However, since the cosmid was identified within a chromosome-1 specific library by hybridisation, it can nonetheless only contain the real *DFFB* gene, which is on chromosome 1. The pseudogene-specific hybridisation probe was still able to detect the real gene because there is sufficient sequence homology between this and the pseudogene (see below) to allow hybridisation at this stringency.

Cosmid 176-g8 was found to contain the entire *DFFB* gene, so was useful in completing the gene structure analysis. PCR across the five introns using cosmid 176-g8 as the template allowed the approximate sizes of these to be determined. Figure 17 shows this analysis of intron size.

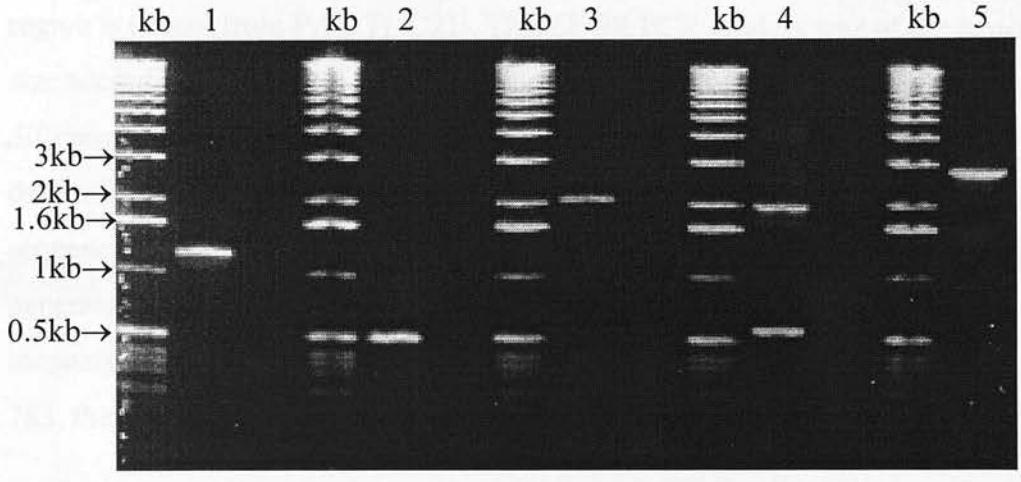


Figure 17. PCRs spanning each intron of *DFFB*, to determine approximate intron size. Lanes 1-5 correspond to PCRs for introns 1-5. Template DNA for all PCRs was cosmid 176-g8. The two bands obtained for the PCR across intron 4 were sequenced, and the smaller band determined to be the result of non-specific annealing of one primer; the larger band represents the intron.

3.2.7 Identification of a *DFFB* pseudogene on chromosome 9

As described above, during the analysis of the genomic structure of *DFFB*, a PCR product of 311p corresponding to the 3' end of the cDNA, was generated from genomic DNA, using primers 6F and 9R. At this stage, analysis of this region of the gene could only be performed on genomic-DNA-derived PCR products, since this region is absent from PAC 71-C21. This 6F/9R PCR product was of the expected size according to the cDNA, but sequencing of this product revealed several differences from the published cDNA sequence, in the form of substitutions, deletions and insertions. In contrast, when cosmid 176-g8 became available, sequencing with primers 6F and 9R directly on this chromosome 1-derived cosmid generated a sequence identical to the published cDNA with none of these inconsistencies, and also revealed a further splice junction between residues 782 and 783, thereby dividing this large region into exons 6 and 7.

The most likely explanation for these results is the existence of a pseudogene for *DFFB* elsewhere in the genome. Pseudogenes are non-functional copies of genes, and can be processed or non-processed. Within these two categories pseudogenes may or may not be expressed. Processed pseudogenes comprise the exonic regions of an active gene, and are thought to arise by the integration into the genome of cDNA sequences generated by reverse transcription of RNA transcripts. Non-processed pseudogenes originate from gene duplication events. In the case of *DFFB*, since at least intron 6 is absent from the pseudogene, it is likely to be processed. Non-functionality is certain, because of the numerous substitution mutations and deletions that appear to have accumulated in the pseudogene sequence, that include frameshifts. A comparison of the sequence differences in this region between the pseudogene and the *DFFB* gene on chromosome 1 is shown in Figure 18.

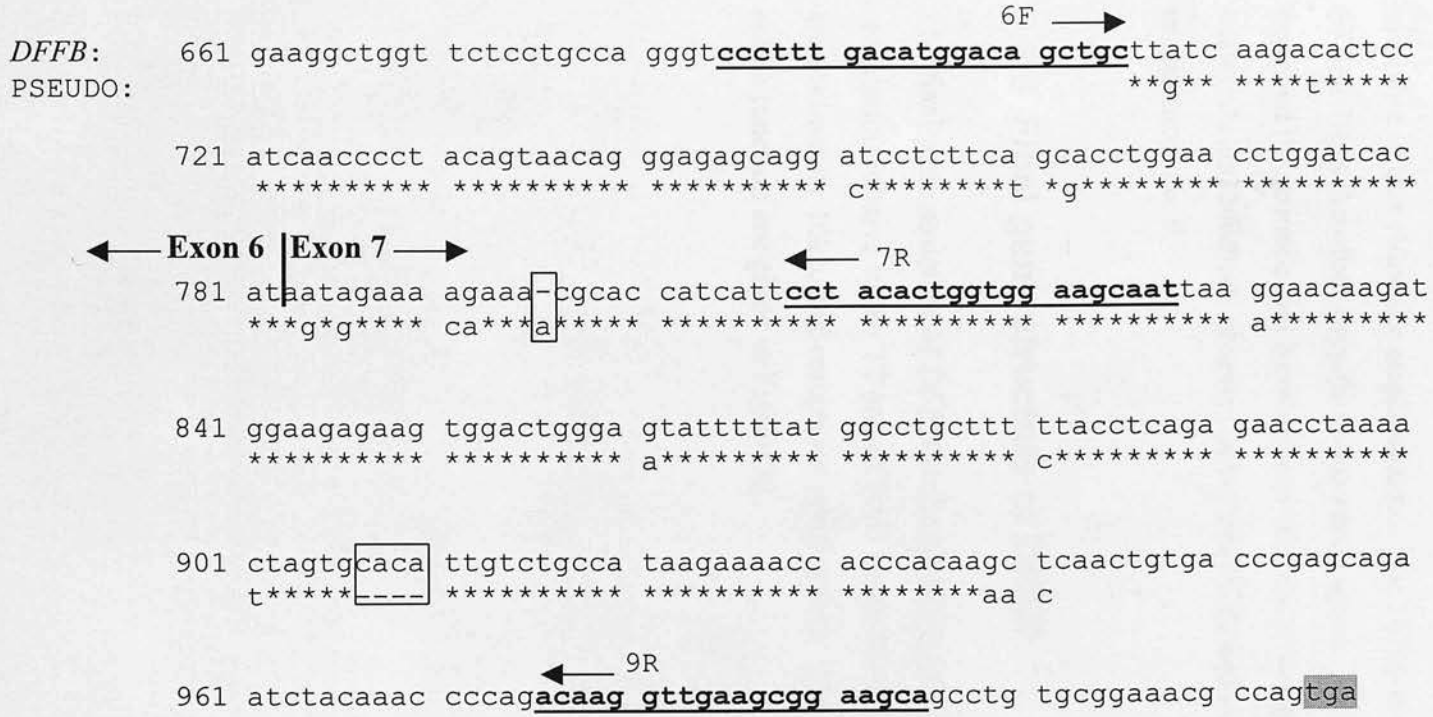


Figure 18. Sequence comparison of the *DFFB* gene on chromosome 1, and the pseudogene on chromosome 9. The numbering is according to the cDNA sequence (Accession AF039210). The upper sequence corresponds to the *DFFB* gene; the lower sequence is the pseudogene. Primers 6F, 7R, and 9R are shown in bold underlined type. An asterisk represents identity between the two sequences. A dash represents a single base deletion. The splice junction between exons 6 and 7 in the *DFFB* gene is shown with a vertical line. The boxes highlight the two positions at which mutations to the pseudogene result in frameshift alterations. The stop codon is shaded.

PCR with primers 6F and 7R (located in exons 6 and 7 respectively) generates a product of 143bp from a genomic DNA template, but a much larger product of approximately 1.4kb from cosmid 176-g8. This is consistent with the pseudogene data above; when genomic DNA is the template, the absence of intron 6 in the pseudogene causes preferential amplification from the pseudogene rather than from the genuine *DFFB* gene, because the product from the pseudogene is smaller. On the other hand, when cosmid 176-g8 is used as the template, only the real gene is available in this clone for amplification. The 143bp product generated with primers 6F and 7R is therefore specific to the pseudogene. Taking advantage of this fact, I screened the somatic cell hybrid panel (details of this panel are provided in the Materials and Methods chapter) using this PCR, and was able to map the pseudogene to chromosome 9.

3.2.8 Final gene structure of *DFFB*

The final gene structure of *DFFB* is shown in Figure 19, along with its location and orientation relative to the T7 end of PAC clone 71-C21. The gene spans approximately 10kb, and comprises seven exons. The sequences adjacent to each splice junction are given in Figure 20.

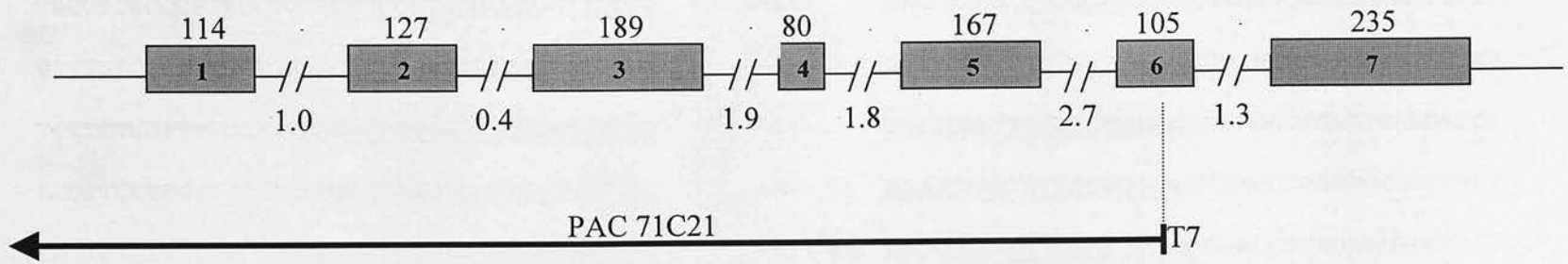


Figure 19. Structure of the *DFFB* gene, and its position relative to the PAC clone 71C21. Grey boxes represent the exons. Upper numbers are the exon sizes in base pairs; lower numbers are the intron sizes in kilobases.

ATGCTCCAGAAGCCCAAGAG	--1--	AGGGCTGTCTCCGCTTCCAGGTGCCCGCTGGGCTAGGCGG
TGCTTCTCCCGTCCCTGCAGCTCCCTGAGCGCGGTTCCCG	--2--	GGGCCAGGCCTGGCAGGGCTGTGAGTGGCAAGGACTTTGG
GTTTGTCCCATTGGTGGCAGATGTGAGCGACATCAGGCGC	--3--	GGACCCGCCGTGGTTTGAAGGTGCGTGGGGGCTGCAGCTG
TCTTCTCGTTTTCCTTGCAGGCTTGGAGTCCCGATTTCAG	--4--	TCCGGAGTTACCTGAGGGAGGTGAGCCTGAGTGAAGACCG
ACGTTCCTTGGTTCCTCCAGGTGAGCTCCTACCCCTCCAC	--5--	AAGGCTGGTTCTCCTGCCAGGTGAGCTGTGTGCCCTTTAT
CATTGTCTTTTGGCCCCCAGGGTCCCTTTGACATGGACAG	--6--	ACCTGGAACCTGGATCACATGTAAGCTCACAGAGCGAGGT
TCTCTAACCTTACTTTGCAGAATAGAAAAGAAACGCACCA	--7--	CTGTGCGGAAACGCCAGTGA

Figure 20. The sequences of the *DFFB* splice junctions. Numbers represent exon number. Underlined sequences are exonic; twenty base pairs of adjacent intron sequence are given for each splice junction.

3.2.9 Mutation screening of *DFFB* in a panel of neuroblastomas

The candidacy of *DFFB* as a neuroblastoma tumour suppressor gene was next addressed. This was performed through rigorous mutation screening of the entire coding region of the gene in a panel of DNAs from neuroblastoma patients, and cell lines, obtained from Dr. F Speleman. This panel comprises 14 neuroblastoma cell lines and 11 primary neuroblastoma tumours. Also included in the panel were DNAs derived from 5 Merkel cell carcinoma cell lines, 10 Merkel cell carcinoma tumours, and 2 Schwannomas. These were included because they possess some biological similarities to neuroblastoma, and have been associated with deletions of 1p (Van Gele *et al.* 1998a; Van Gele *et al.* 1998b). In total, the panel consists of 42 samples. A summary of the properties of the DNAs in this panel is shown in Tables 2 and 3.

Tumour/Cell Line No.	Tumour/Cell Line Code	1p LOH	D1Z2 copy	D1Z1 copy	MNA
Neuroblastomas					
Cell lines					
1	IMR32	Yes	1	3	Yes
2	N206	Yes	2	3	Yes
3	UHG-NP	Yes	2	4	Yes
4	TR14	Yes	2	3	Yes
5	KCNR	Yes	1	2	Yes
6	STA-NB-3	ND	2	4	Yes
7	STA-NB-8	ND	1	2	Yes
8	STA-NB-9	ND	1	2	Yes
9	STA-NB-10	ND	1	2	Yes
10	STA-NB-12	ND	1	2	Yes
11	NGP	t(1;15)	2	2	Yes
12	SK-N-AS	Yes	1	3	No
13	GI-ME-N	Yes	2	4	No
14	SK-N-FI	No	2	2	No
Primary Tumours					
15	g3(EM, st1)	ND	ND	ND	No
16	g11(VDBS, st4S)	No	ND	ND	No
17	g20(VRT, st4)	No	ND	ND	No
18	g22(AY, st4)	No	ND	ND	No
19	g29(VHM, st4)	No	ND	ND	No
20	g25(RM, st4)	Yes	ND	ND	Yes
21	g1(DAM, st1)	ND	ND	ND	ND
22	g2(DSLT, st1)	ND	ND	ND	ND
23	g8(PS, st2)	ND	ND	ND	ND
24	g9(PB, st2)	ND	ND	ND	ND
25	g10(VPS, st2)	ND	ND	ND	ND

Table 2. Details of Neuroblastoma tumours and cell lines used for mutation screening of *DFFB*. MNA = *MYCN* amplification; ND = not determined. 1p status was determined by LOH, and in cell lines was also assessed by FISH with the probes D1Z2 at 1p36.33 and D1Z1 at 1q12 (decreased copy number of D1Z2 relative to D1Z1 indicates a deletion).

Tumour/Cell Line No.	Tumour/Cell Line Code	1p LOH	D1Z2 copy	D1Z1 copy	MNA
Merkel Cell Carcinomas					
Cell lines					
26	MCC 13	No	2	2	No
27	MCC 14/2	t(X;1)	4	4	No
28	MCC 26	Yes	3	4	No
29	UISO	ins 1p36	2	2	No
30	MKL-1	ND-	2	2	No
Tumours					
31	UGH-VM	Yes	ND	ND	No
32	UGH-RM	No	ND	ND	No
33	UGH-FA	No	ND	ND	ND
34	RJ	No	ND	ND	ND
35	MCC 2T	t(1;5)	ND	ND	Yes
36	K2967	Yes	ND	ND	ND
37	K1213	No	ND	ND	ND
38	K681	No	ND	ND	ND
39	W8179	No	ND	ND	ND
40	W6388	No	ND	ND	ND
Schwannomas					
41	PNST AA	Yes	ND	ND	No
42	PNST BC	Yes	ND	ND	No

Table 3. Details of Merkel cell carcinoma tumours and cell lines, and schwannomas included in the *DFFB* mutation screen.

Pairs of intronic primers were designed to flank each exon of *DFFB*, so that in each case, approximately 50-100bp of intron was included each side of the exon amplicon. This was to ensure that when sequencing the PCR products, the entire exon and splice junctions would be analysed. These primer pairs are named cpan-ex1-pcr5'/cpan-ex1-pcr3', cpan-ex2-pcr5'/cpan-ex2-pcr3' etc (see 2.A.1). Each of the seven exons was amplified from each of the 42 DNAs in the panel, and the PCR products checked on 1% agarose gels before sequencing radioactively on both strands. Sequencing primers were designed internally to the PCR primers, but still within the intronic sequence. These primers are named cpan-ex1-seq5', cpan-ex1-seq3' etc. To simplify the analysis, the sequences were run on the gel in groups of approximately six samples, with all 'A's run alongside each other, then all 'C's, then all 'G's, and finally all 'T's.

In the 42 samples screened, only three nucleotide substitutions were detected in the coding region of *DFFB*, and seven nucleotide substitutions were detected in intronic sequences. However, all of these were detected also in normal DNA samples, suggesting that they are common sequence polymorphisms rather than tumour-specific alterations. These are detailed in table 4. Of the three exonic polymorphisms, two are silent, and the third results in a conservative amino-acid substitution of arginine 196 for lysine. The functional effect of this substitution on the resulting protein is unknown. However, this polymorphism is not tumour-specific, so it is unlikely that it plays any causative role in neuroblastoma. The insignificance of both this polymorphism, and the two silent polymorphisms with respect to neuroblastoma, is substantiated by their occurrence in normal DNA controls, as well as neuroblastomas. Figure 21 shows the detection of the two exon 5 polymorphisms, and Figure 22 shows the exon 7 polymorphism.

Position	Nature of polymorphism		Allele frequencies in normal controls	
Exon 5, residue 21	G → A	Silent	G = 18/30	A = 12/30
Exon 5, residue 77	G → A	Conservative substitution: Lysine → Arginine	G = 27/30	A = 3/30
Exon 7, residue 172	G → A	Silent	G = 58/108	A = 50/108
5' utr, -38	A → G		N.D.	
5' utr, -30	T → C		N.D.	
IVS 1, +15	A → G		N.D.	
IVS 1, +31	T → G		N.D.	
IVS 3, +44	T → C		N.D.	
IVS 3, -72	G → A		N.D.	
IVS 3, -82	T → C		N.D.	

Table 4. Details of polymorphisms within the *DFFB* gene, identified during mutation screening. N.D. = Not Determined. IVS = Intervening Sequence. NB. The positions given for the 5' utr are relative to the start codon, whereas positions for the other intervening sequences are relative to the splice junctions.

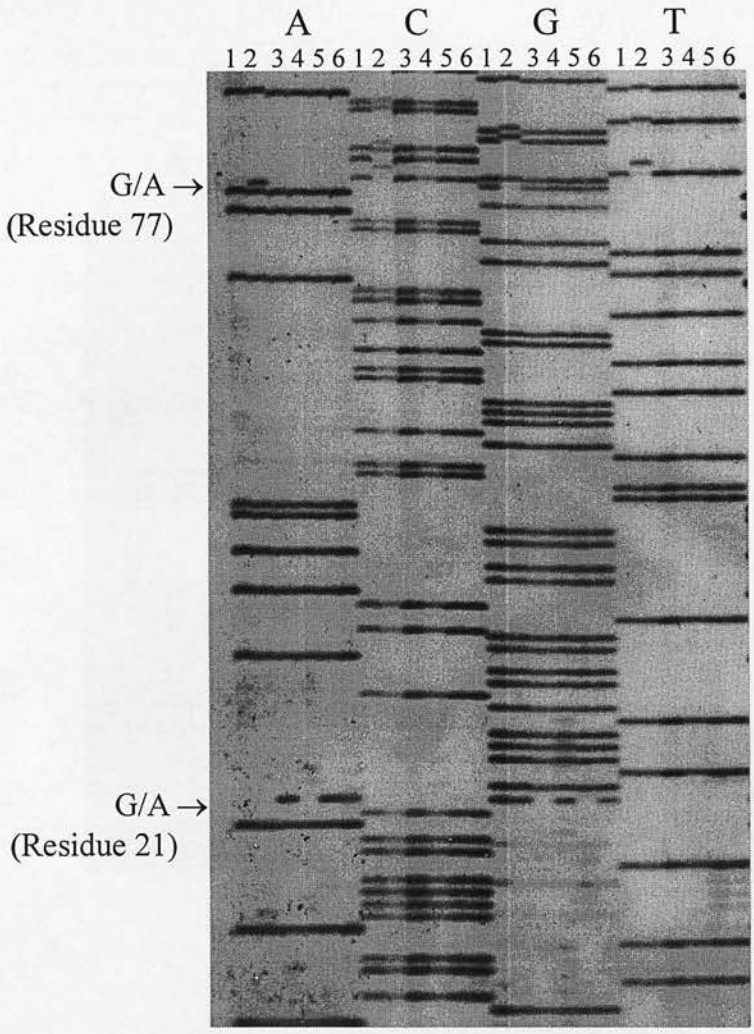


Figure 21. Example of sequencing gel showing detection of polymorphisms during *DFFB* mutation screening of neuroblastoma DNAs. Forward sequence from exon 5 in six individuals (1-6) is shown, and the G/A polymorphisms at residues 21 and 77 of exon 5 can be seen. At position 21, individuals 3 and 5 are either homozygous or hemizygous for the A allele (depending on whether they have a deletion on one chromosome), whereas individual 6 has both A and G alleles and is therefore heterozygous at this locus. At position 77, individual 2 is either homozygous or hemizygous for the A allele. NB. All As are run in adjacent lanes, followed by all Cs, etc, for ease of detection.

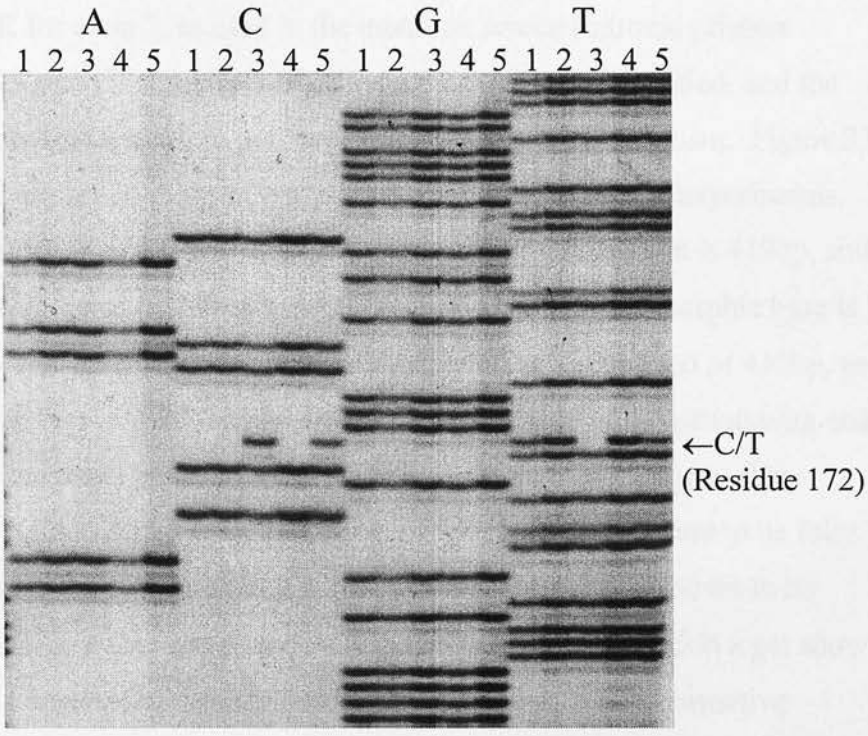


Figure 22. Example of sequencing gel showing detection of polymorphisms during *DFFB* mutation screening of neuroblastoma DNAs. Reverse strand sequencing of exon 7 in five individuals (1-5) is shown, which identifies a C/T polymorphism (G/A on forward strand) at residue 172 of exon 7. Individual 5 possesses both the C and T alleles, and is therefore heterozygous for the polymorphism, whereas individual 3 is either homozygous or hemizygous for the C allele only.

3.2.10 Is *DFFB* imprinted?

Three exonic polymorphisms and seven intronic polymorphisms were identified during the neuroblastoma mutation screen. The most common polymorphism identified, at residue 172 of exon 7 alters a site for the restriction enzyme *Ava*I. Using the PCR for exon 7, as used in the mutation screen (intronic primers CPANex7pcr-5' and CPANex7pcr-3'), the fetal DNAs were amplified, and the products digested with *Ava*I, to genotype them for this polymorphism. Figure 23 is a diagram of the design of the genotyping and subsequent RT-PCR experiments, showing the location of the primers. The genotyping PCR product is 410bp, and is digested into fragments of 270bp and 140bp only when the polymorphic base is 'G' rather than 'A'. A heterozygote will therefore show the uncut band of 410bp, as well as the other two bands of 270bp and 140bp. An example of the genotyping analysis performed on the panel of fetal DNAs is shown in Figure 24. Twenty-three heterozygous (G/A) fetuses were identified, of which six were found to be fully informative since the corresponding maternal DNA sample was shown to be homozygous (G/G in four cases, and A/A in two cases). Figure 25 is a gel showing the genotyping analysis of these maternal DNA samples. The informative heterozygous fetuses were 9' (MC260166), 17' (MC151165), 23' (HW300462), G (NG100369), L (JR170474), and M (FC070271).

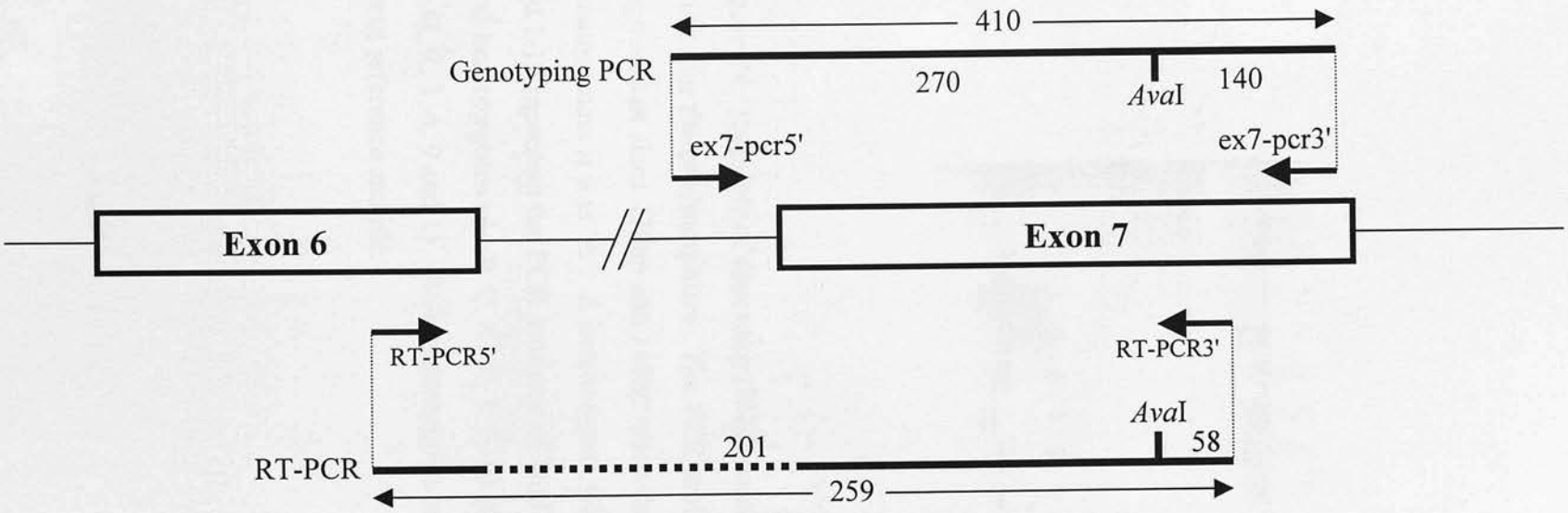


Figure 23. Schematic diagram to show the location of the primers and the polymorphic *Ava*I site in the genotyping PCR and RT-PCR experiments for the *DFFB* exon 7 polymorphism. The sizes of the products and digestion fragments are given in base pairs. Not to scale.

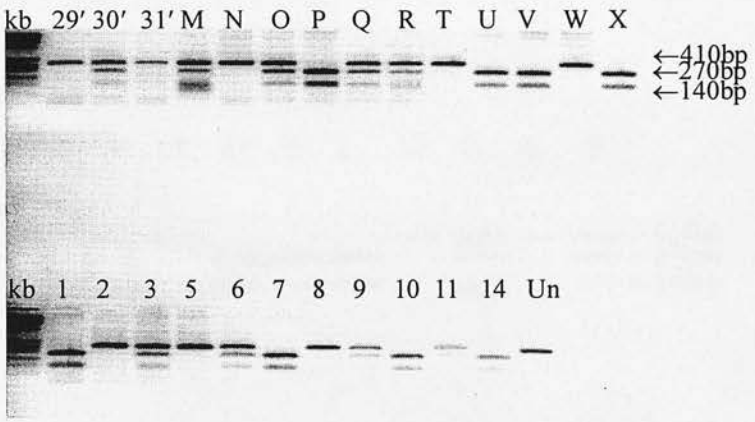


Figure 24. Example of digesting *DFFB* exon 7 PCR products with *Ava*I to genotype fetuses for the polymorphism. The PCR product of 410bp is digested into two fragments of sizes 270bp and 140bp when the polymorphic residue is 'G', but remains uncut if it is 'A'. A heterozygote will show all three bands. Lanes 29' – X and 1-14 represent the PCR products of fetal DNAs, digested with *Ava*I. On this gel, G/G homozygotes are: P, U, V, X, 1, 7, 10 and 14. G/A heterozygotes are: 30', M, O, Q, R, 3, 6, 9 and 11. A/A homozygotes are 29', 31', N, T, W, 2, 5 and 8. Un is an uncut reference sample.

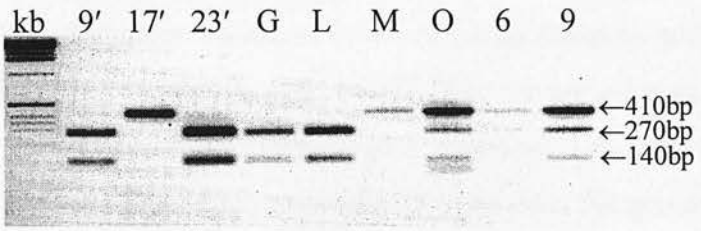


Figure 25. *Ava*I-digested maternal exon 7 PCRs corresponding to heterozygous fetuses. Maternal DNAs homozygous for this polymorphism are: 9', 17', 23', G, L and M. The corresponding fetal DNAs of these samples are therefore fully informative.

An RT-PCR experiment was designed and performed on RNA from fetuses 9' (MC260166) and 17' (MC151165). The primers for this RT-PCR, named CPANRT-PCR5' and CPANRT-PCR3', span the splice junction between exons 6 and 7, to give a product of 259bp. This is digested into bands of 201bp and 58bp when the polymorphic *Ava*I site is present.

Before analysing the parental origin of the RT-PCR products, an experiment was designed to ensure that the RT-PCR product was derived only from *DFFB*, and not from the pseudogene identified on chromosome 9, since this could potentially have distorted the results. Sequence variations between the pseudogene and the genuine *DFFB* gene remove a site for the enzyme *Alw*44I from the pseudogene, so a simple digest of the RT-PCR products with this enzyme, which showed complete digestion, confirmed that the RT-PCR product was solely derived from the genuine message.

Digestion of the RT-PCR products with *Ava*I revealed bands of 259bp and 201bp in all samples, indicating that both allele types are present, and *DFFB* is therefore biallelically expressed. The digests were separated on an 8% polyacrylamide gel, which is shown in Figure 26. The 58bp band is not visible on this gel.

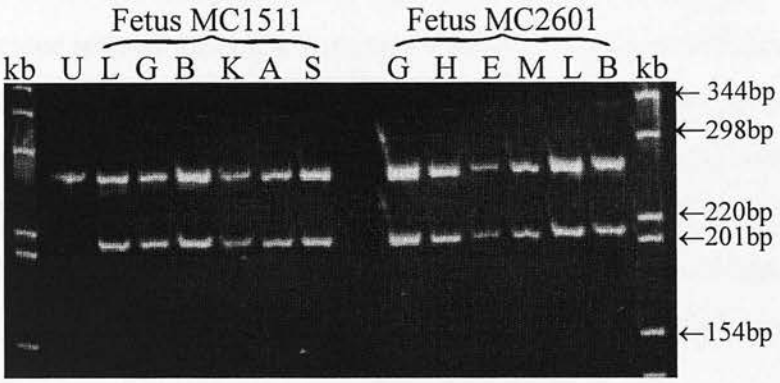


Figure 26. *Ava*I-digested RT-PCR products from fetuses MC1511 and MC2601, for *DFFB* exon 7 polymorphism. U is an uncut reference sample of the RT-PCR product, L = lung, G = gut, B = brain, K = kidney, A = adrenal, S = stomach, H = heart, E = eye, M = muscle. The marker is the Gibco BRL 1kb ladder, with band sizes as indicated on the right. Note that the 58bp band that results from digestion of the 259bp fragment is not resolved on this gel.

3.3 DISCUSSION

The results of attempts to determine the expression status of the recently identified gene *TP73* have so far proved contradictory. The initial report of its identification was accompanied by evidence for monoallelic expression (Kaghad *et al.* 1997), and it was perhaps the combination of several significant factors such as possessing tumour suppressor activity and high sequence similarity to p53, in addition to being monoallelically expressed, that provoked such a great interest in this gene. Subsequent research, however, has yielded results that are conflicting with regard to whether *TP73* is monoallelically or biallelically expressed, and with a substantial number of published papers now providing evidence on both sides of this controversy, it seems likely that the imprinting of *TP73* is not a simple story.

The results of experiments presented here indicate that *TP73* is biallelically expressed in fetal tissues. However, as suggested by Nomoto *et al.* (Nomoto *et al.* 1998), it may be the case that *TP73* is imprinted in a tissue-specific and possibly time-specific manner, and perhaps imprinting is variable even between individuals. A classic example of a gene that is imprinted in a polymorphic manner such as this is *IGF2R* (Xu *et al.* 1993; Smrzka *et al.* 1995; and see Chapter 1). Therefore, although the RT-PCR experiments in this thesis show no evidence for imprinting, *TP73* may still be subject to imprinting in some tissues, or in tissues of post-fetal stage. The rationale behind using *TP73* as a starting point for finding imprinted genes at 1p36 was based on the known phenomenon of imprinted genes occurring in clusters. Despite finding no evidence here that *TP73* is imprinted, other studies have demonstrated monoallelic expression, and unless it can be proved that *TP73* is <u>not</u> imprinted, this rationale will remain valid. However, the many studies that have found no evidence for mutations of the *TP73* gene in neuroblastoma, and the overexpression rather than the loss of this gene indicate that *TP73* is unlikely to play a tumour suppressor role in neuroblastoma. It would seem therefore that a tumour suppressor gene remains unidentified as yet in this region.

A contig of PAC clones and a cosmid clone was constructed around the *TP73* locus, and it was by BLAST searching with the end sequences of these clones that part of the sequence of the recently discovered gene *DFFB* (DNA fragmentation factor) was found to reside close to *TP73*. *DFFB*, also know as *CAD* (caspase-activated DNase) (Sakahira *et al.* 1998) or *CPAN* (caspase activated nuclease), encodes a 40kDa apoptotic nuclease, capable of degrading DNA to induce apoptosis (Halenbeck *et al.* 1998). The protein exists in a heterodimeric complex with its inhibitor ICAD/DFF45, and proteolytic cleavage of this inhibitor by activated caspase-3 releases DFFB in its active state. A possible mechanism by which *DFFB* could function as a tumour suppressor gene is apparent in its role in inducing apoptosis. A loss of this function from the cell would naturally result in decreased apoptosis, which could lead to tumorigenesis.

From the published cDNA sequence of *DFFB*, the complete gene structure was deduced. Given the apoptotic role of *DFFB*, and the genomic location, it seemed an excellent candidate for the putative neuroblastoma suppressor gene. A thorough and extensive mutation screen was performed on the entire coding sequence of *DFFB*, in a panel of 42 neuroblastoma DNAs and cell lines. This panel deliberately included both 1p36-deleted and non-deleted tumours. This was to take into account that there are two possible hypotheses regarding the action of the tumour suppressor gene, depending on whether or not it is imprinted. Firstly, if the tumour suppressor gene at 1p36 is imprinted, as predicted, then there is only one functional copy of the gene. Therefore, in deletion cases, if it is always the expressed allele that is deleted, then there need not be any point mutations in the remaining non-expressed allele. However, in non-deleted cases of neuroblastoma, a heterozygous point mutation or other inactivating mechanism might be identifiable as the 'second hit' in the functional copy of the gene, (the first 'hit' being the imprint itself), in order for the gene to undergo complete loss of tumour suppressor function. By the second hypothesis, if the tumour suppressor gene is not imprinted, a 'second hit' mutation would be expected in all deletion cases. Disappointingly, no mutations were detected in any of the tumours analysed.

Imprinting analysis was performed by RT-PCR, but the result suggested that the gene is biallelically expressed. Biallelic expression, together with a complete absence of any coding region mutations, now makes it extremely unlikely that *DFFB* is involved in neuroblastoma. However, it cannot be ruled out that *DFFB* may play a role in other tumour types. Deletion and LOH of this chromosomal region has been implicated in many different cancers, and there may be several tumour suppressor genes associated with 1p36.

In summary, the expression status of *TP73* has been analysed. This gene has been found to show biallelic expression, and no evidence of differential methylation at nearby CpG islands. This gene is therefore unlikely to act as the putative imprinted neuroblastoma tumour suppressor expected at 1p36. A search in the surrounding region has identified *DFFB* as the nearest gene to *TP73* and a possible candidate for the suppressor due to its genomic location and known apoptotic role. In this study the gene structure of *DFFB* has been established, and a thorough mutation screen performed on an extensive panel of neuroblastoma DNAs. No mutations were detected, and *DFFB* was found to be biallelically expressed, thereby making it unlikely that it is the putative imprinted neuroblastoma suppressor gene.

Neither *TP73* nor *DFFB* is likely to function as the neuroblastoma suppressor gene, and the genuine suppressor therefore remains to be identified. A very recent study identified a 500kb homozygously deleted region at 1p36.2-36.3 in a neuroblastoma cell line, thus narrowing down the candidate interval expected to contain the suppressor gene (Ohira *et al.* 2000). Sequencing across this interval has so far identified six genes that may be considered as candidates for the putative suppressor gene, including *DFF45/ICAD*, *HDNB1/UFD2*, *KIAA0591F* (Nagase *et al.* 1998), *PEX14* (Will *et al.* 1999), *PGD* (Tsui *et al.* 1996), and *CORT* (Ejeskar *et al.* 2000). It has been demonstrated by Ohira *et al.* that expression of the first four of these is high in favourable neuroblastomas but low in unfavourable neuroblastomas. They also demonstrated that mutations of these genes were infrequent in neuroblastoma, but whether any of the genes are subject to imprinting (which acts as a first Knudson 'hit' and thus may preclude the identification of mutations in deletion cases) has not

yet been addressed. In the case of *DFF45/ICAD*, since this is the gene encoding the inhibitor of *DFFB*, it is difficult to envisage how a loss of this function would have any tumorigenic effect, as this would more likely lead to an increase in apoptosis given that *DFFB* would be uninhibited. However, *in vitro* experiments by Zhang *et al*. (Zhang *et al.* 1998a) demonstrated that the cells of mice carrying a mutant *DFF45* gene had no DFF40 nuclease activity, and were resistant to DNA fragmentation and chromatin condensation, suggesting that both subunits are needed to produce the DNase activity. The loss of either subunit therefore might be considered to have a potential role in tumorigenesis. Future studies should be directed at identifying further candidates in this region, and analysing them in a similar way for both monoallelic expression and tumour-specific coding region mutations.

CHAPTER 4

ZAC, AN IMPRINTED CANDIDATE GENE FOR TRANSIENT NEONATAL DIABETES MELLITUS (TNDM)

4.1 INTRODUCTION

Following the first discovery of the imprinted genes, *Igf2* and *Igf2r*, in 1991, there are now approximately thirty known genes that show an imprinting effect. The identification of several of these genes has been due largely to chance, but the development of systematic screening methods for detecting imprinted genes has always been regarded as the most effective way to find new imprinted genes.

The knowledge that imprinted genes are generally associated with differential methylation of CpG islands, and differential expression of the two parental alleles, immediately provides two key criteria that could form the basis of systematic screening methods. Indeed, four successful methods have been developed from these starting points: mRNA differential display and cDNA subtractive hybridisation, which are based on differential expression; RLGS (restriction landmark genome scanning) and RDA (representational difference analysis), which are both based on differential methylation.

As reviewed by Kelsey and Reik, (Kelsey, Reik 1998), the limiting factor in the application of any of these screening techniques is the availability of appropriate material for comparative analysis. In the case of the mouse, all four techniques have been successfully used to identify imprinted genes. This has been possible both because of the opportunities that exist for embryo manipulation in mice and because of the ability to control the genetic background. In the first place, it is possible to generate embryos that are completely uniparental in their genetic background; parthenogenetic embryos are produced by the spontaneous or artificial activation of oocytes, whereas androgenetic embryos are constructed by pronuclear transplantation techniques. Material from these uniparental embryos has proved suitable for the identification of imprinted genes by cDNA subtractive hybridisation screening (Kaneko-Ishino *et al.* 1995). A second resource that is of great value in the identification of imprinted genes is mice with uniparental disomy for a specific chromosome or chromosomal segment. A third resource is mice that are hybrids of two different parental strains, in which the parental alleles can be distinguished.

Analysis of imprinted genes in this case relies on detecting sequence differences, particularly at the RNA level, using RT-PCR, or ribonuclease protection assays. The advantage of hybrid mice over the other artificially generated resources is that they do not suffer the phenotypic effects of an imbalance in their imprinted gene complement, and material can be obtained from all developmental stages.

These controllable genetic backgrounds are all readily available in mice. Much greater difficulties, however, arise in applying similar screening methods to the identification of human imprinted genes, where ethical constraints limit the availability of suitable material for comparison by these techniques. One biological resource which could be exploited using these screening methods for the detection of human imprinted genes is DNA from the parthenogenetic chimera patient FD (Strain *et al.* 1995). The leukocytes of this patient are exclusively maternal in origin, and therefore represent a human parthenogenetic DNA resource which is highly useful for comparison with normal DNA, or with complete hydatidiform mole, a tumour derived from androgenetic extra-embryonic tissue (Kajii, Ohama 1977) which is the closest, reciprocally equivalent resource to the FD DNA. For both RLGS and RDA, the parthenogenetic human cell line derived by EBV transformation of this patient's lymphocytes should be a significant resource for detecting differential methylation between maternal and paternal genomes.

4.1.1 Restriction landmark genome scanning (RLGS)

Restriction landmark genome scanning (RLGS) relies on the use of rare-cutting methylation-sensitive restriction enzyme sites, such as *Not*I as 'landmarks' for CpG islands. The technique involves digesting the target DNAs with a methylation-sensitive enzyme, end labelling the fragments with ^{32}P, and then digesting with a second enzyme, before resolving the fragments in one dimension by electrophoresis. A further digestion is then performed *in situ* with another more frequently-cutting enzyme, and the gel electrophoresed in a second dimension. The resulting autoradiograph reveals up to two thousand spots, each of which represents a site for the methylation-sensitive enzyme used in the initial digestion. Where differential

methylation of such a site occurs, the methylated allele will not be cleaved, and therefore will not give rise to a spot on the final autoradiograph. Conversely, the unmethylated allele will give rise to a spot signal, hence the ability of this method to detect differential methylation.

In the laboratory of Y. Hayashizaki, the RLGS technique was used to compare the parthenogenetic DNA from FD, with normal DNA, and DNA from androgenetic hydatidiform mole, the main genetic difference between these samples being their exclusive maternal or exclusive paternal origin respectively. The enzyme combinations used were *Not*I-*Pvu*II-*Pst*I, *Not*I-*Pst*I-*Pvu*II, and *Asc*I-*Eco*RV-*Mbo*I. Two spots were identified as being differentially methylated (and in both cases maternally methylated) by their absence from FD DNA, relative half intensities in normal DNA, and full intensities in androgenetic hydatidiform mole DNA. These two spots were named A20 (identified from the enzyme combination *Asc*I-*Eco*RV-*Mbo*I), and NV149 (identified from the enzyme combination *Not*I-*Pst*I-*Pvu*II). Cloning and subsequent investigation of the A20 spot clone revealed that this was associated with the imprinted gene *GNAS1* (Hayward *et al.* 1998a).

The purpose of this study was to investigate the RLGS spot clone NV149, identified from the RLGS screen described above, on the basis of a differentially methylated *Not*I site. As described below, my initial mapping placed NV149 on chromosome 6, and subsequently refined this location to 6q24. The chromosomal region 6q24 is very interesting, not least because of its frequent association with a wide range of tumour types, but also because it is now known to carry a paternally expressed imprinted gene responsible for transient neonatal diabetes mellitus, or TNDM. Each of these aspects is discussed further below.

4.1.2 Transient neonatal diabetes mellitus (TNDM)

Transient neonatal diabetes mellitus (TNDM) occurs in approximately 1 in 500 000 live births (von Muhlendahl, Herkenhoff 1995). It was first described in 1852, and accounts for 50 to 60% of neonatal diabetes. The condition presents with

intrauterine growth retardation, hyperglycaemia, a general inability to thrive, and in some cases dehydration (Milner *et al.* 1971; Shield *et al.* 1997). There is always inadequate insulin production, which is either very low, or undetectable, and the patients require exogenous insulin therapy for the first few months of life. The median time of insulin therapy required is 3 months. TNDM presents in the first six weeks of life, but resolves spontaneously within the first year, and insulin treatment becomes unnecessary after this time. However, despite its transient nature, TNDM often predisposes to insulin-dependent diabetes later in life, often within childhood (Shield *et al.* 1997; Von Muhlendahl, Herkenhoff 1995). There is no evidence of anti-islet antibodies or type 1-associated HLA class II susceptibility haplotypes, and it has been suggested that the condition is related to a defect in β-cell maturation (Shield *et al.* 1997; Ferguson, Milner 1970).

4.1.3 The genetic basis of TNDM

The majority of TNDM cases are sporadic, but in one third of reported cases, familial recurrences have been observed, although the inheritance patterns do not follow the usual Mendelian laws (Cave *et al.* 2000). Evidence for a genetic basis to TNDM began to accumulate from the diagnosis of affected siblings, cousins, and half-siblings in the late 1960s and 1970s; (Coffey, Womack 1967; Ferguson,Milner 1970; Milner *et al.* 1971). However, despite this early knowledge of an underlying genetic basis to TNDM, it was not until 1995 when it was first proposed that an imprinted gene might be involved in TNDM. Temple *et al.* (Temple *et al.* 1995) described two TNDM patients, both of whom showed paternal uniparental disomy (UPD) for chromosome 6, suggesting that paternal UPD, and either the resultant increased dosage of a paternally expressed imprinted gene, or the resultant lack of a maternally expressed gene, plays a significant role in TNDM. This report was followed up by further studies from the same group, of two unrelated families, both of which provided additional, and independent evidence that the underlying genetic cause of TNDM is related to a paternally expressed imprinted gene mapping to 6q22-23 (Temple *et al.* 1996). In their study, Temple *et al.* investigated two families with affected TNDM individuals. The first family (family A) included a female child

with TNDM born to unaffected parents. Using PCR analysis of dinucleotide repeat polymorphic markers in the 6q22-23 region, a duplication was detected in the affected child, her father, and her paternal grandmother. The duplication was detected by the presence of either three (rather than two) alleles of a marker, or increased dosage of one allele relative to the other. The experimental method used here enabled a critical region to be determined - there was no evidence of duplication at markers D6S472 or D6S311, suggesting therefore that these markers flank a region containing the putative TNDM gene. See Figure 27. This was the first study to precisely define a critical TNDM region, although the defined region itself was still relatively large, spanning approximately 60cR, or 16Mb, according to the LDB map. This family was particularly useful in gaining an understanding of the genetics of TNDM, since it not only allowed the establishment of a critical region, but also confirmed what was already suspected with regard to the imprinted status of the putative gene. The child had inherited the duplication paternally, and was affected, yet her father who carried the same duplication was unaffected, having inherited it through the maternal inheritance line. His mother was also unaffected, and it can be assumed that she also must have inherited the duplication maternally. In brief, the affected child inherited two paternal copies and one maternal copy of the 6q22-23 region, whereas the unaffected father inherited two maternal copies with one paternal copy. By logical deduction therefore, it is the inheritance of two paternal copies of this region that gives rise to TNDM. The inheritance pattern in this family substantiates the proposal that the TNDM gene is imprinted, and expressed only from the paternal allele.

The second family (family B) investigated by Temple *et al.* (Temple *et al.* 1996) had previously been reported by Ferguson and Milner (Ferguson, Milner 1970), Milner (Milner *et al.* 1971) and Wilson (Wilson 1991). The four affected individuals in this family, in two different generations, appeared karyotypically normal, and molecular techniques revealed no apparent duplication, deletion or uniparental disomy. However, positive linkage of TNDM with marker D6S310 at 6q23.3 has been shown in this family, with a lod score of 2.7. The inheritance pattern also shows an

imprinting effect, since the linked D6S310 allele is consistently paternal in origin in all affected family members.

The main conclusions from the study by Temple *et al.* (Temple *et al.* 1996) of families A and B, are that TNDM is caused by the up-regulation and resultant increased expression of a paternally expressed imprinted gene, which resides in a 60cR critical region of chromosome 6q flanked by markers D6S472 and D6S311.

Clarification of an important point is needed here: how can we be sure that TNDM results from up-regulation of a paternally expressed imprinted gene? It could be argued that in TNDM cases with paternal uniparental disomy of chromosome 6, it is impossible to distinguish whether it is the presence of two paternal copies of chromosome 6 (and the resultant double dose of paternal genes), or the effect of the maternal loss of chromosome 6, that is responsible for TNDM. The answer lies in two crucial observations: firstly, paternal duplications of the 6q region give rise to TNDM in the same way as paternal UPD6, although in duplication cases the maternal chromosome is still present, so the common factor between paternal UPD6 cases and paternal 6q duplication cases is the extra paternal copy of the region, rather than maternal loss. Secondly, there have been several reported cases of maternal deletions of the 6q region, yet none are associated with TNDM (Pandya *et al.* 1995).

The fact that the reciprocal situation of maternal UPD6 or maternal duplications of the equivalent 6q region do not cause TNDM (Pivnick *et al.* 1990), emphasises further that the disease is due to the effect of a paternally expressed imprinted gene.

In order to make progress in identifying the putative TNDM gene, it was essential to refine the critical region, by identifying and studying new patients, as well as further characterising the duplications in existing patients. In the next report, by Arthur *et al.* (Arthur *et al.* 1997), the critical TNDM region defined by Temple *et al.* (Temple *et al.* 1996), was further reinforced, following their diagnosis of TNDM in a girl born to unaffected Ashkenazi Jewish parents. Karyotypic investigation of the patient revealed an inverted duplication of 6q22-23, and subsequent molecular investigation

with polymorphic markers defined the region as spanning at least 10cM. The duplication in this case includes markers D6S270, D6S314, D6S1684, D6S310, and D6S308, and covers a very similar region to that described by Temple *et al.* However, uninformative markers in the family prevented the extremes of the duplication being accurately defined in this case.

The first genuine refinement of the critical TNDM region came from Gardner *et al.* (Gardner *et al.* 1999). Using new information from families A and B, which they had characterised previously (Temple *et al.* 1996), they were able to refine the locus to an 18cR (5.4Mb or 3-4cM) region, between markers D6S1699 and D6S1010. The result which led to this refinement was a combination of data from FISH experiments which further defined the duplication in family A, and from a comprehensive microsatellite analysis from 14 members of family B which allowed the identification of a TNDM haplotype in this family.

Another significant refinement of the critical TNDM region came from a new, large study by Cave *et al.* (Cave *et al.* 2000). This group performed microsatellite analysis of DNA from 13 affected individuals in 9 families. In particular, comparative mapping of the duplications found in three different families allowed a shortest region of overlap to be determined, which is flanked by markers D6S308 and D6S1010. These markers are known to be separated by a distance of less than 1cM, so this significantly reduces the critical region. Figure 27 illustrates the narrowing down of the critical TNDM region by each of the studies described above.

Having defined this small region of 6q as the critical region containing the putative TNDM gene, identification and characterisation of candidate genes within this region is the obvious choice for future progress. Several genes mapping to 6q have been implicated in diabetes, including *PDNP1*, encoding a transmembrane glycoprotein (Buckley *et al.* 1990), and the putative diabetes susceptibility loci *IDDM15* (6q21) (Delepine *et al.* 1997), *IDDM5* (6q24-27) (Davies *et al.* 1994) and *IDDM8* (6q25-27) (Luo *et al.* 1996), but these can now be eliminated from any involvement in TNDM because they do not map to the defined region between D6S308 and D6S1010. The

identification of a gene mapping to the correct location would be a valid candidate for involvement in TNDM, particularly if it was found to be imprinted, and had an appropriate biological role.

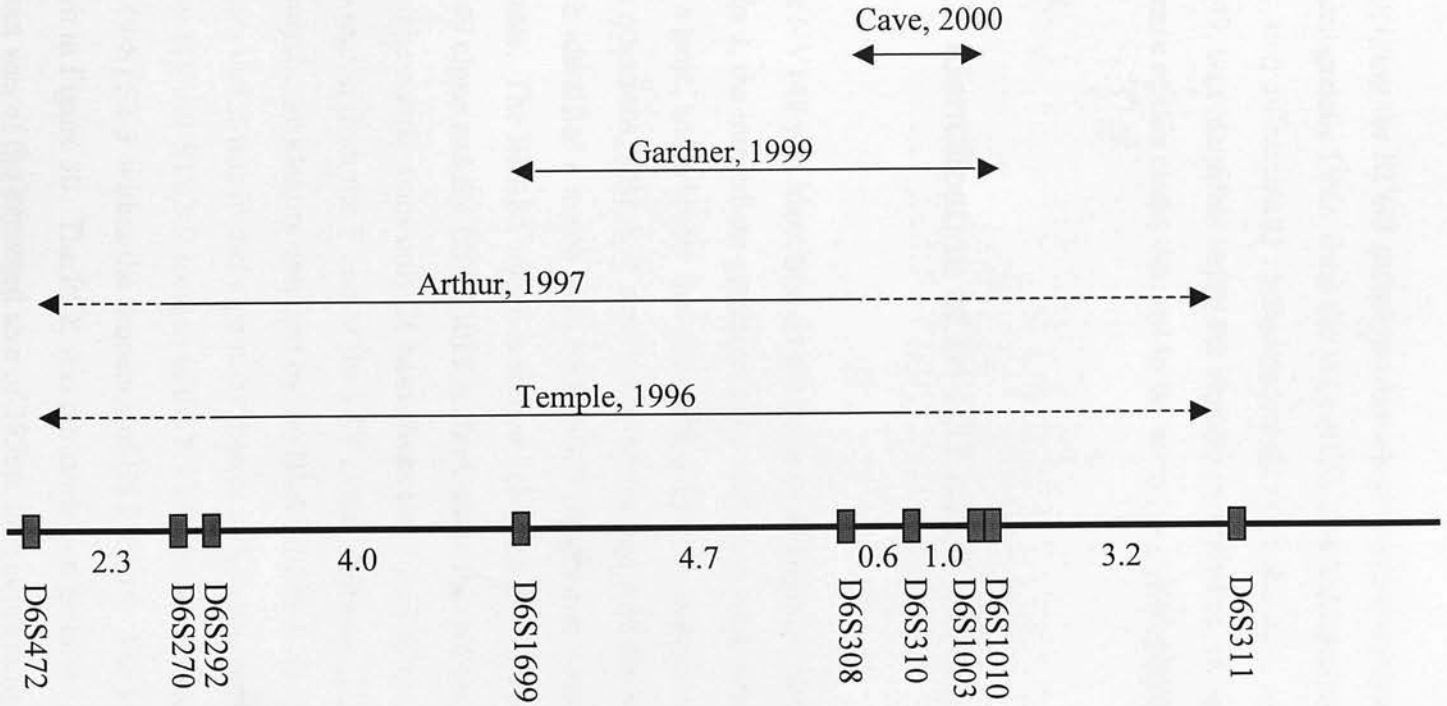


Figure 27. The candidate TNDM intervals as defined by the studies of Temple, 1996; Arthur, 1997; Gardner, 1999; and Cave, 2000. The map is derived from the genetic location database LDB, and the distances given are in Mb. Markers within the intervals shown by solid lines are part of the duplications in patients, whereas the dotted lines indicate the potential extent of the interval given that not all markers were analysed, and some were non-informative. Markers outside the indicated intervals were not duplicated, and therefore define the limits in each case. See text for more information.

4.2 RESULTS

4.2.1 Identification of NV149 through RLGS screen

By applying the RLGS technique described above to compare normal DNA with parthenogenetic DNA from the patient FD, and androgenetic hydatidiform mole DNA, two differentially methylated spots were detected. One of these, named NV149, was identified using the enzyme combination of *Not*I-*Pst*I-*Pvu*II. The sequence of this clone, derived in the laboratory of Y. Hayashizaki, is given in Figure 28.

4.2.2 Identification of an EST clone overlapping the NV149 clone

Since NV149 was identified on the basis of differential methylation of the *Not*I site within it, the immediate questions to be answered were whether it was associated with a gene, and whether that gene was subject to imprinting. To begin to answer these questions, a BLAST search was performed with the NV149 sequence. This search identified a match with EST 53019 (Accession number AA346811) in the database. The BLAST match is shown in Figure 29. The identity between the NV149 clone and the EST 53019 is short, since the two sequences overlap at one end. The match spans only 21 bases from the 3' end of the NV149 sequence with the same region from the 5' end of the EST clone. Because of the very short length of this match, the identity detected by the BLAST search appears weak. However, to confirm that it was in fact a genuine match, a PCR was designed with the forward primer NV149-STS2-5' located in the NV149 sequence, and the reverse primer NV149-STS2-3' within the sequence of EST 53019. The location of these primers is shown in Figure 30. The PCR was performed on genomic DNA, and the resulting product was of the expected size of 295bp, thus confirming that the sequence within the EST clone 53019 is contiguous with the genomic sequence of NV149, and is therefore a genuine match.

***Not*I**

GCGGCCGCGAGGAGGGTGTGCCTTTGCCGCGCCGCCTACGTGCGGGTCCGGGCTCCG

CGGGGCCGGGTGCGGGACCCCGCAGATCGTCACCCGCAACCCAGGCAGCCCCACCGC

GAGTGCCGCCGGACCCCCTGGACGCCGCTGCCAGAGGCGTTCGCGCCTATCTGGTAT

GAGGTCCACAGACCCGATTCTTACAACCTGGCGCTCTAACCTCGCCAACGGGCCAGG

AAAAAAACAGAACAAGGAAAAGAAGAAAAAGTCTGTTCCAAGTAATAATGGGACTAG

ACAGTAACTGTTTGCACTTTCGTCTCTTATGGAAAATATGATTATTTTGATGCTTTA

GTATTACAGACTGTATCAGACATACTCTTAAAAGTATTTAAAAACGTGGGGTCCTAA

AAATTTCCTAGATCGGGA**CAG(ctg)**

***Pvu*II**

Figure 28. The sequence of the differentially methylated spot NV149, detected by RLGS.

```
gb|AA346811.1|AA346811 EST53019 Fetal heart II Homo
sapiens cDNA 5' end.
           Length = 236

 Score = 42.1 bits (21), Expect = 0.055
 Identities = 21/21 (100%)
 Strand = Plus / Plus



Query: 400 aaatttcctagatcgggacag 420
           |||||||||||||||||||||
Sbjct: 1   aaatttcctagatcgggacag 21
```

Figure 29. Result of a BLAST search of NV149 against the EST database. The final 21 bases of the NV149 sequence have complete identity with the first 21 bases of an EST clone 53019.

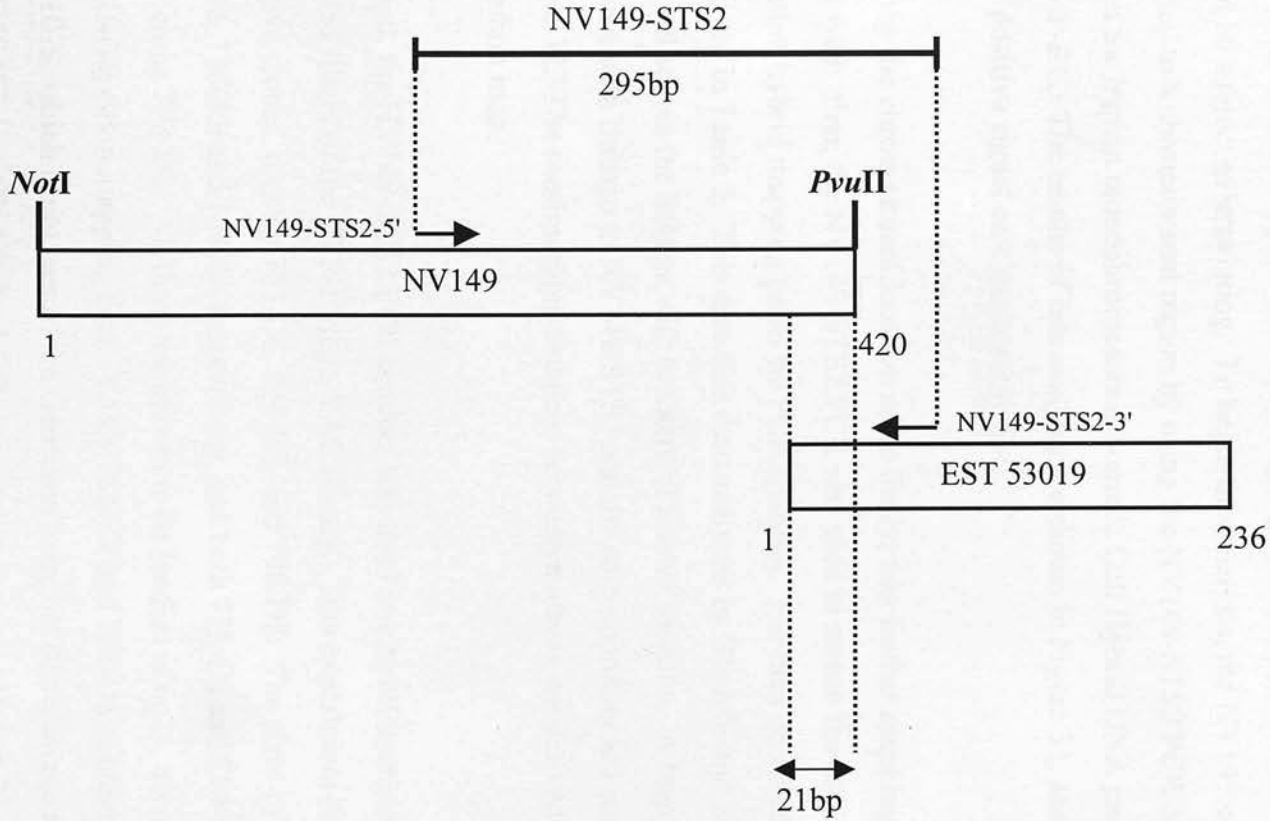


Figure 30. Diagram showing the 21bp overlap between the NV149 sequence and EST clone 53019, and the position of PCR primers for NV149-STS2.

4.2.3 Localisation of NV149 to 6q24

The NV149 sequence, now extended by the overlapping sequence of EST 53019, represents a differentially methylated CpG island, since it was identified on the basis of differential methylation of the *Not*I site at the 5' end of the sequence. CpG islands are almost always associated with genes, and it was therefore of interest to pursue the NV149 locus and analyse the surrounding region for the presence of genes which might be subject to imprinting. To begin this analysis, the NV149 clone was first mapped to a chromosomal region by using the NV149-STS2 PCR described above to screen the human monochromosomal Somatic Cell Hybrid DNA panel from the UK HGMP-RC. The results of this analysis are shown in Figure 31, and NV149 gives a clear positive signal on chromosome 6.

To map the chromosomal location more finely, two further mapping techniques were employed: first, the NV149-STS2 PCR was used to screen the Genebridge 4 radiation hybrid mapping panel by PCR analysis. The data obtained from this screen are given in Table 5. This data was then analysed by RhyME and Sanger systems to give lod scores for linkage with markers of known location. A map showing the markers with linkage to NV149-STS2 and the corresponding lod scores is given in Figure 32. The centimorgan distances between markers are derived from the Généthon map.

Second, the NV149-STS2 PCR product was used as a hybridisation probe for the gridded filters of the CEPH Mega-YAC library. This experiment detected three positive clones, namely 733-C6, 774-H5, and 798-D8. The sizes of these YACs are 730kb, 1360kb and 1740kb respectively, and both 733-C6 and 774-H5 overlap the third clone 798-D8. All three are known to be located at 6q24, according to the CEPH-Généthon mapping data. YACs 733-C6 and 798-D8 contain the marker D6S1079, which is not part of the Généthon map, but lies between markers D6S978 and D6S453 according to the LDB map, and is shown in Figure 32.

Together, the mapping data from the Somatic Cell Hybrid panel, the radiation hybrid mapping panel, and the CEPH Mega YAC library indicate that NV149 resides at 6q24, with linkage at the highest lod score of 9.347 to polymorphic microsatellite marker D6S978.

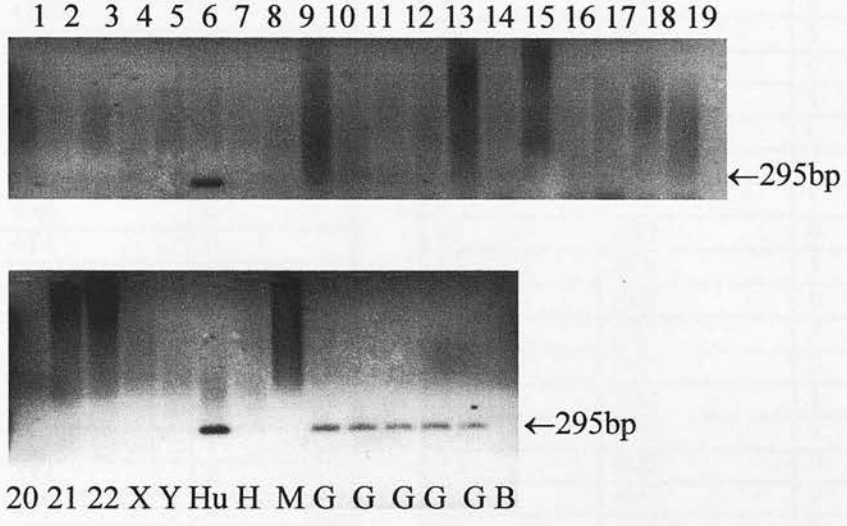


Figure 31. Somatic cell hybrid panel mapping of NV149-STS2 PCR to chromosome 6. The panel consists of chromosomes 1-22, X, Y, Human (Hu), Hamster (H), and Mouse (M) controls. Five additional human DNAs (G) are included; B=Negative control PCR. The 295bp product is present only in the chromosome 6 hybrid, and human controls.

Table 5. Raw data obtained from PCR analysis of Radiation Hybrid mapping panel.

0 = negative result; 1 = positive result; 2 = unknown/ambiguous result.

Tube Number	Cell Line	PCR Result
1	4G6	0
2	4B3	0
3	4E6	0
4	4F13	0
5	4P9	0
6	4E2	0
7	4Z11	0
8	4Z9	0
9	4BB6	0
10	4D1	0
11	4E4	0
12	4Y9	0
13	4BB1	1
14	4BB12	0
15	4N6	0
16	4T4	0
17	4U1	0
18	4N12	1
19	4AA7	1
20	4C11	0
21	4T10	0
22	4DD2	1
23	4Z5	0
24	4S3	0
25	4I4	0
26	4N3	0
27	4O5	0
28	4L6	0
29	4T11	0
30	4A5	0
31	4R1	0
32	4C3	1
33	4M4	0
34	4W1	0
35	4G7	0
36	4L4	0
37	4N7	0
38	4I1	0
39	4R6	0
40	4BB10	0
41	4N5	0
42	4G11	0
43	4H12	0
44	4U3	0
45	4G1	0
46	4DD8	2
47	4H1	2
48	4J5	0
49	4P11	0
50	4B9	0
51	4K5	0
52	4O10	0
53	4H9	1
54	4J9	1
55	4M5	1
56	4F7	1
57	4Y4	0
58	4CC8	0
59	4Z6	0
60	4K7	1
61	4B2	0
62	4R10	1
63	4Z12	0
64	4AA5	1
65	4T3	1
66	4D7	0
67	4Q4	1
68	4P2	0
69	4R12	0
70	4K9	0
71	4A4	1
72	4Y8	1
73	4Q2	0
74	4H8	0
75	4DD5	0
76	4S12	1
77	4F6	0
78	4L3	1
79	4S10	1
80	4R2	1
81	4V2	2
82	4J2	1
83	4R3	0
84	4V7	0
85	4G5	1
86	4V8	0
87	4R5	0
88	4BB8	0
89	4K8	0
90	4S6	0
91	4K12	0
92	4V3	0
93	4E11	1
94	A23 (Hamster control DNA)	0
95	HFL (Human control DNA)	1

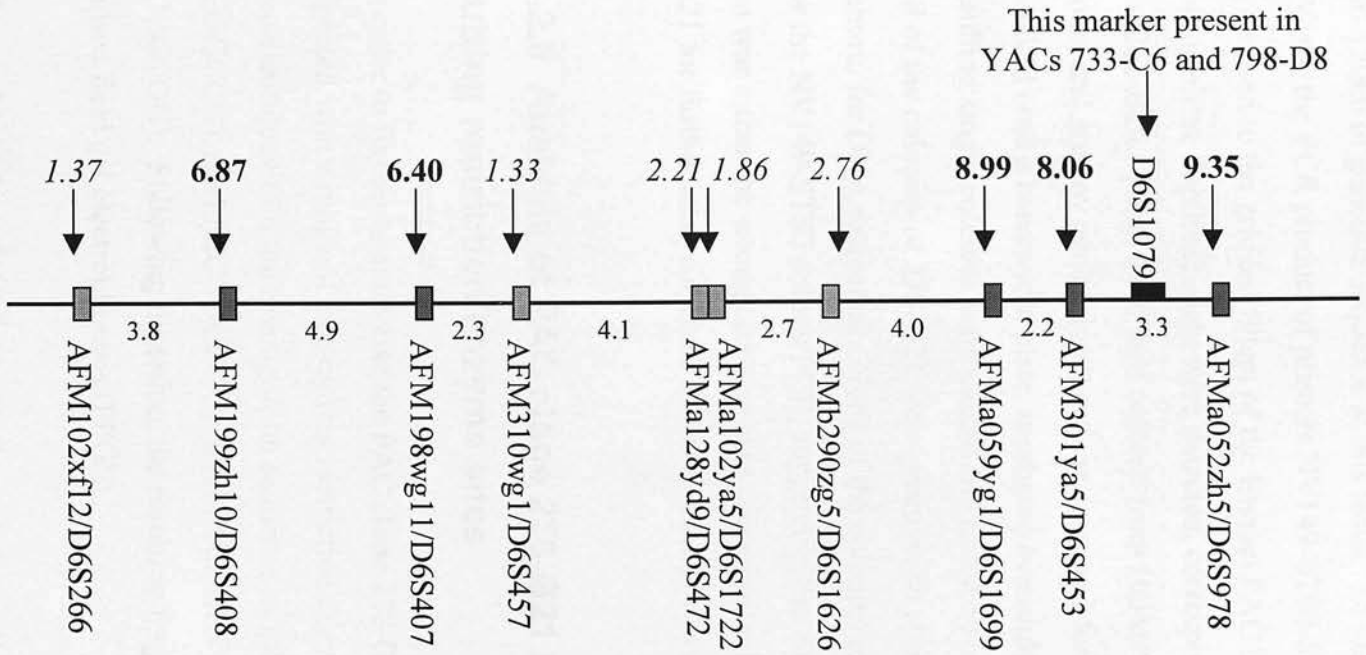


Figure 32. Representation of the markers showing linkage to NV149-STS2, based on results of radiation hybrid mapping. The map is derived from the Généthon human genetic linkage map, and the numbers below the line are sex averaged distances between markers in cM. Lod scores are given above the line; those from the RhyME analysis are in bold type, and those from the Sanger analysis are in italic type. Based on these data, the highest lod score is for the marker AFMa052zh5/D6S978. Marker D6S1079, which lies between D6S978 and D6S453, is present in two of the three overlapping YAC clones that are positive for NV149-STS2.

4.2.4 Identification of a PAC clone containing NV149

In order to investigate and characterise the region surrounding the NV149 CpG island, it was decided to obtain a PAC clone, which would provide approximately 100-150kb of genomic sequence at this locus. To identify a PAC clone containing NV149, the PCR product of primers NV149-STS2-5'/3' was labelled with ^{32}P, and hybridised to the gridded filters of the Human PAC library RPCI1 from HGMP-RC. Two positive duplicate spots were detected, corresponding to PAC clones 278-G21, and 262-B20. These were then ordered from HGMP-RC. To ensure that the correct clone, and not any contaminating clone was used for further analysis, the clone was streaked onto a kanamycin plate, incubated overnight, and approximately ten of the resulting single colonies were picked and tested by colony PCR for NV149-STS2. All of the colonies of 278-G21 were positive, so one was picked and grown as a large culture, for DNA extraction. None of the colonies of 262-B20 gave a positive signal for the NV149-STS2 colony PCR, suggesting that the clone that had been streaked out was either the wrong clone, or was contaminated. It was decided to pursue 278-G21 for further analysis, since it contains the NV149 putative CpG island.

4.2.5 Analysis of PAC clone 278-G21 for the presence of rare-cutting restriction enzyme sites

In order to further characterise the PAC clone 278-G21, DNA was prepared, and then digested with a range of rare-cutting restriction enzymes containing a CpG dinucleotide within their recognition sequence as follows: *Bss*HII (GCGCGC), *Eag*I (CGGCCG), *Not*I (GCGGCCGC), *Mlu*I (ACGCGT), *Pvu*I (CGATCG), *Sac*II (CCGCGG). Following digestion, the resulting fragments were electrophoresed by pulsed field gel electrophoresis (PFGE).

The results of the pulsed field gel of the *Not*I and *Eag*I digests are shown in Figure 33. The *Not*I digest generated fragments of approximately 60kb, 50kb, and a vector fragment of 16kb (since the insert is cloned into a *Not*I site within the vector). *Eag*I

(CGGCCG), being contained within the recognition site of *Not*I (GCGGCCGC) always cuts at the same sites as *Not*I, and in this case the *Eag*I digestion generated the same 50kb and 16kb fragments that were observed in the *Not*I digest. It also generates fragments of approximately 40kb and 20kb, suggesting that there is an additional *Eag*I site within the 60kb *Not*I fragment.

The restriction enzymes *Bss*HII and *Sac*II both generated fragments of approximately 20kb, 40kb and 65kb, whereas neither *Mlu*I nor *Pvu*I appear to have recognition sites within this PAC clone sequence (data not shown).

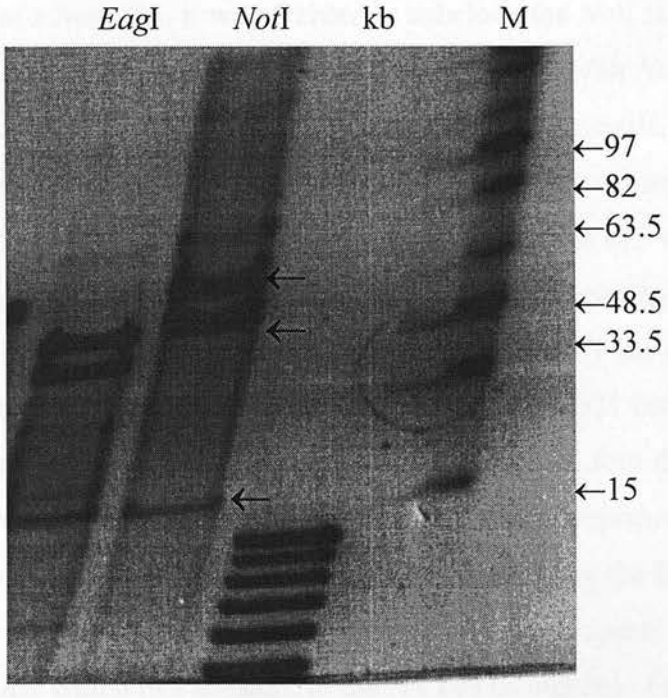


Figure 33. Pulsed field gel electrophoresis of PAC clone 278-G21 digested with *Not*I and *Eag*I. M=Mid range marker (NEB), marker band sizes are shown at the side in kb. The arrows indicate the bands resulting from the *Not*I digest, which are approximately 50kb, and 60kb in size; the 16kb band represents the vector.

4.2.6 Subcloning the *Not*I sites in PAC clone 278-G21

The results of the rare-cutting restriction enzyme analysis described above indicated that PAC clone 278-G21 contains one internal *Not*I site, which cuts the insert into fragments of approximately 60kb and 50kb. Based on this result, because the NV149 clone lies adjacent to a *Not*I site, it was decided to subclone the *Not*I site of this PAC clone. DNA from PAC 278-G21 was digested simultaneously with *Not*I and *Sac*I, and the resulting fragments were ligated into vector pCRII (for details, please see Materials and Methods), which was then used to transform electrocompetent JM109 cells. Of the resulting white colonies, 18 were picked and grown in 5 ml cultures overnight, for miniprep. The inserts were digested out of the vector by a double digest with *Not*I and *Sac*I, and run on a 1% agarose gel to identify the different classes of subclone. From the knowledge that PAC clone 278-G21 contains one *Not*I site, in addition to the artificial *Not*I sites which flank the insert, four different types of subclone are expected from this experiment: two subclones representing the T7 and SP6 ends of the PAC clones, and two subclones representing the internal *Not*I site detected in the PFGE experiment (including one which represents the NV149 sequence, and another which lies adjacent to the NV149 sequence). Figure 34a is a schematic diagram to explain this.

It was apparent from digesting out the inserts that several classes of insert size were present among the 18 subclones. In order to distinguish between these, each subclone was sequenced using the M13 forward primer to sequence out of the vector, across the *Not*I site, and into the insert. Approximately 200bp of sequence were obtained in each case; the beginning of each sequence is given in Table 6. Unexpectedly, eight different classes of subclone insert were detected. Each sequence was then compared to the database by BLAST searching, and this revealed that one insert was in fact a sequence derived from *E. coli*, and another represented a subclone with no insert. However, excluding these two subclones still leaves six different insert types instead of the expected four.

Subclones 1, 3, 7, 13 and 16 represent the expected NV149 subclone class. By sequencing directly on PAC 278-G21 with a primer in this NV149 subclone sequence, pointing back towards the *Not*I site, it was possible to determine that subclone 2 represents the adjacent subclone, sharing the same NV149 *Not*I site. Sequencing of the T7 and SP6 ends of the PAC clone 278-G21 indicated that subclone 12 is the SP6 end subclone, and subclones 6, 9 and 11 represent the T7 end subclone. However, the remaining two subclone classes (ie, subclone 10; subclones 14, 15, 17, 18) remain novel, and surplus to the expected four subclone classes. The most feasible explanation for this is that PAC clone 278-G21 contains another *Not*I site, which was not detected in the PFGE experiment. This is quite possible, particularly if the two *Not*I sites are relatively close to each other, because in addition to the two large fragments that were seen, a *Not*I digestion would also generate a very small fragment not visible at the resolution level of PFGE.

Subsequent to this experiment, sequencing at the Sanger centre chromosome 6 project produced complete sequences of several PAC clones at this locus. One of these, PAC clone 340-H11, shares a substantial overlap with PAC clone 278-G21, and the sequence data from this clone revealed the above prediction to be correct, because there is an additional *Not*I site only 420bp away from the NV149 *Not*I site, with a *Sac*I site between the two. This explains the additonal two *Not*I-*Sac*I subclones generated in this study. Figure 34b shows the actual arrangement of PAC clone 278-G21 with respect to the *Not*I and *Sac*I sites.

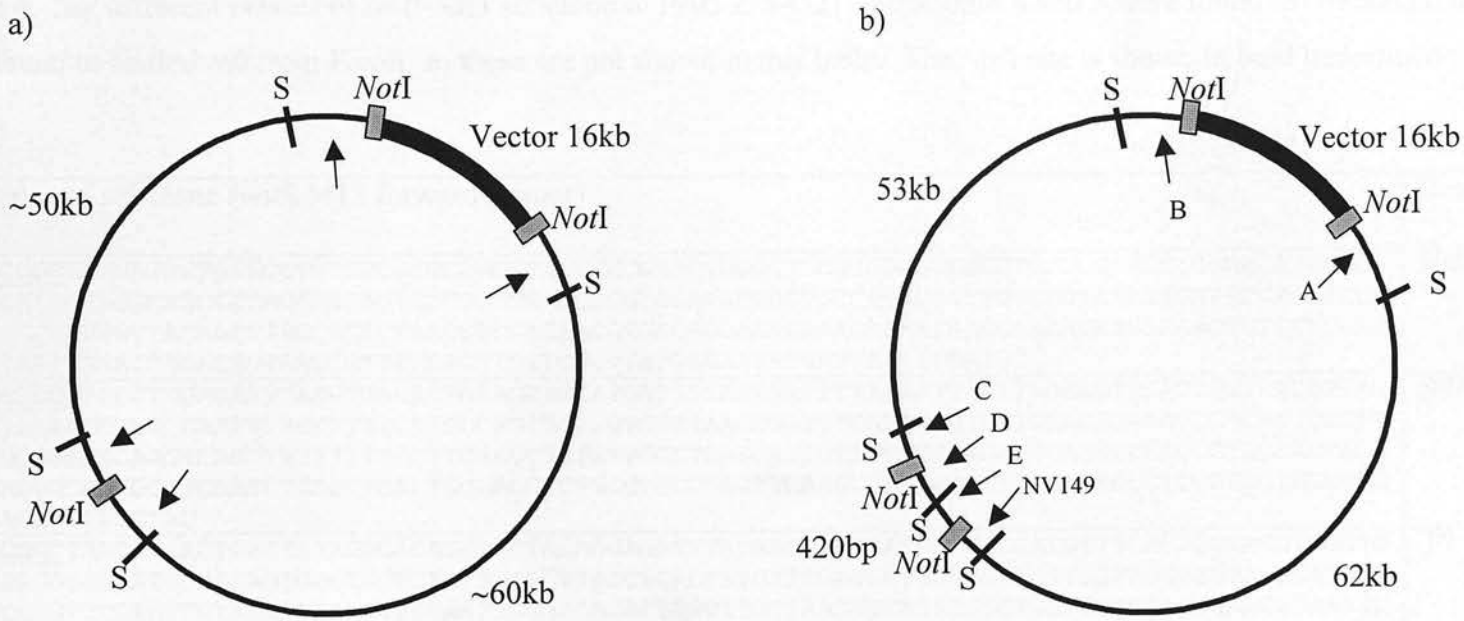


Figure 34. a) Expected *Not*I-*Sac*I subclones of PAC clone 278-G21, based on the results of the pulsed field electrophoresis of digest fragments. *Not*I=*Not*I site, S=*Sac*I site. Approximate estimated distances between *Not*I sites are given. The four expected subclones are indicated with arrows. b) Actual restriction map based on published PAC clone sequences. The presence of an additional *Not*I site 420bp from the NV149 *Not*I site explains the existence of six different classes of subclone, rather than the predicted four. Actual distances between *Not*I sites are shown in kb or bp, as appropriate. Subclones A and B are the T7 and SP6 end subclones; C and D are the subclones associated with the additional *Not*I site; E is the subclone sharing the same *Not*I site as NV149. Not to scale.

Table 6. Six different classes of *Not*I-*Sac*I subclone of PAC 278-G21. Subclones 4 and 5 were found to contain no insert, and subclone 8 was found to be derived from E.coli, so these are not shown in this table. The *Not*I site is shown in bold underlined type.

Sequence of subclone (with M13 forward primer)	Description	Subclones of this type
GCGGCCGCGAGGAGGGTGTGCCTTTGCCGCGCCGCCTACGTGCGGTCCGGGCTCCGCGGGGCCGGGTGCGGGACCCCCCGAGATCGTCA CCCGCAACCCAGGCAGCCCCACCGCGAGTGCCGCCGGACCCCCTGGACGCCGCTGCCAGAGCGTTCGCGCTATACTGGTATGAGGTCCA CAGACCCGATTCTTACAACCTGGCGCTCTAACCTCGCCAACGGGCCAGGAAAAAAACAGAACAAGGAAAAGAAGAAAAGTCTGTTCCAA GTAATAATGGGACTAGACAGTAACTGTTTGCACTTCGTCTCTTATGGAAATATGATTAATTTGATGC	NV149	1, 3, 7, 13, 16
GCGGCCGCCTGGCTCGACATTTAGGTGACACTATAGAAGGATCATGTCAGTGATTTTAGAATTATTTAGAGGTTGATTGCTCACATTAG GACCTCCAACCTTGCCAAATCATCTTTTCTTGCCTGTTGTTGTCTCTAACTGGTTTCATTATATGCTGTACTGGGACCCACATTCTGTT TTCTGGTTAAGCAAGTCAGTTATTTTTTCCTTCAACCTCTGTTGCCTCATCATCCTTTTTTTGAGATGCAGTCCCACCCTGTCACCAGG CTGAAGTGCAGTGGCTCGATGTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGCGATCCTCCTACCTCAGCCTCCCGAGTAGCTGG GATTACAGGCGTGTAC	SP6	12
GCGGCCGCTAATACGACTCACTATAGGGAGAGGATCTACAGGAGACTTATGAGGATTAACAGCTACGATGTTTCAGAGGCACCTGGCTG TGAAAAGGCCACATTCAGGAGTGGCTGCTCTCTTGTTCTTACCTCACTTTGTTTGATCCTGGTGGAACTTGGTAGGTTGCACAGATAGA CTGTCCTGTTCCAGTTGTAGAAAGAAATTGGAATTGGAAGAGACTGACTTGCTAACGTCATATGGCTGATGGGTAACTGAAGTGAAATG AGAGGCTCCCTTTTGGAATTTAGTCCTGCGTGTTTCTTCTGCTCCATGCTATGGTTGGCTCTCTTCTGGCTGTTGTGACTCTCCCTGGA GTGAAAAGGAGCAAAGGCTGTAATCATGCATGTCTGTACAA	T7	6, 9, 11
GCGGCCGCAGACCCAGCGCCGTCCTCGCAGCGCGGCAATGCACGGCCACCGCTGCCCCCAGCCCGCCCGCCGCAGCCGCGAGCACCCAA ACACCTACCCTGCGGGCGACGACCCGGAGCTCCTAGAGCAAAAATTGGAGAAACTTTTTGTTATTTGGTTCTTAAATTATTTTAAGT	Adjacent to NV149	2
GCGGCCGCGCGCTTCCGGCTCCCGAGCCCCGCCTGCGCGCGGCCTCCTCGGCGCACGATCCTCTTGGCTGCCGCGGGCGCGAAAGCCCA CGGCATCTGCCATTTGTCATTCAGCCCGTCGGTACCGCCCCGAGCCTGATTTAGACACGGCTGGGGCGTGCTCTGGCCTCACTCTCCGG GCGGGTGCTGGACGGACGGACGGACGGGGCAGCCGTGCTCACAGCTCAGCAGCGCGGGCCTTGGCGCCGGGGCGTTCCCCGGGTCGCGT CATGGCCGC	?	10
GCGGCCGCGAGGAGGGCGCTGGGGCCCCTGGCGGGGCGGTCACGTGGCAGGAGGAGGCCCCGCCGGGAGCTGGGGGTCGGCGGCCGAGG CGGGGGAGCTGAGCGCACCCACACGTCCTGCGGGCCGGGTCACCGTGGGGGCAAAGCAAGGTCTCCCAGGTACAGCGCTGGCGAGGTAG ACCCGAGCCGGCCTGGGGTCTGCAGCGGGGCCTGCTAGCCGAAGTCTCCGCCAGATGGGCCGCCAGAGCAATCACACATGAGAAACGCG ACAGA	?	14, 15, 17, 18

4.2.7 Identification of IMAGE clone 2073154 (AI540783) at the NV149 locus

The M13 forward sequence of the NV149 subclone (1, 3, 7, 13 and 16), together with that of the adjacent subclone (2), gives a contiguous sequence of 643 nucleotides around the *Not*I site of NV149. A BLAST search with this sequence identified a human cDNA IMAGE clone, 2073154 (AI540783). Figure 35 shows the position of this clone. The IMAGE clone 2073154 contains an insert of 1388bp, which I sequenced in full.

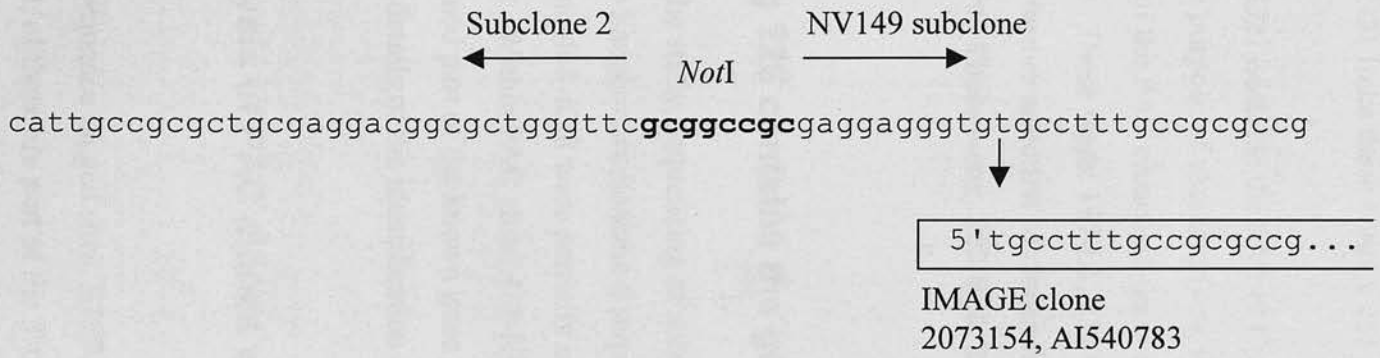


Figure 35. The position and orientation of the 5' end of the IMAGE clone 2073154 (AI540783), relative to the *Not*I site of NV149, and the two associated 278G21 subclones. The *Not*I site is shown in bold type.

4.2.8 PAC clone 278-G21 is part of Sanger contig 226

Sequencing at the T7 and SP6 ends of PAC 278-G21 generated sequences which, when compared with the Sanger databases, revealed that this PAC clone is part of the Sanger contig 226. Additionally, the T7 sequence of 278-G21 matched part of the PAC clone 197-L1, which itself was part of a different Sanger contig, 318. Thus, PAC clone 278-G21 links these contigs 226 and 318 together.

Since PAC 278-G21 contains the NV149 CpG island, the two contigs 226 and 318 are useful for the purpose of characterising the surrounding region. In order to do this, a selection of the PAC clones on these contigs was obtained from the HGMP Resource Centre. These were: 197-L1, 210-B1, 218-N6, 24-J7, 80-G20 and 96-M11. Subsequently, a further selection of clones from the two contigs was obtained from the Sanger Centre. These were: 262-B20, 357-B1, 430-G4, 468-K18 and 1155-M13.

4.2.9 Contig 226 contains the gene *ZAC/LOT1/PLAGL1*

At this point in the study, sequencing of some of the clones in these contigs was underway by the Sanger chromosome 6 sequencing project. In particular, PAC clones 197-L1 and 468-K18 were partially sequenced. Features of 197-L1 included two CpG islands. Within PAC clone 468-K18, the feature of most interest was that this clone contained part of the known gene *ZAC/LOT1/PLAGL1*. A summary of the background and details of the identification of this gene is given below in 4.2.11.

4.2.10 Analysis of PAC clones within the Sanger contigs 226 and 318

Two published sequence tagged sites, STS9563 and STS24967 were present on contig 226. Both of these are part of the 3' utr of *PLAGL1* and *LOT1*, and it was decided to screen the HGMP PAC clones for their presence (the Sanger clones had

not been obtained at this point). The result of this is shown in Figure 36, which indicates that both STSs are present only in PAC clone 80-G20. All other tested clones were negative; 197-L1, 210-B1, 218-N6, 24-J7 and 96-M11.

The PAC clones were characterised further by screening with another PCR, designed to amplify part of the coding region of *ZAC* (primers hzacpoly1-5'/2237-3'). This PCR was present in PAC clones 80-G20, 262-B20, and 357-B1, as shown in Figure 37. This PCR was negative on PAC clone 468-K18, which was an unexpected result since this part of *ZAC* (including the region amplified by this PCR) is known to be present in 468-K18. It was suspected that the clone 468-K18 was in fact an unrelated clone sent by mistake. This seems to be the correct explanation, as in addition to lacking the *ZAC* PCR described above, sequencing at the ends of this clone generated sequences which did not match the T7 and SP6 end sequences in the Sanger database. This clone was therefore excluded from further analysis. The PAC clones containing at least part of *ZAC*, as determined by this analysis, or from the Sanger sequencing data are 80-G20, 262-B20, 357-B1, and 468-K18.

Figure 36. PCR screening of PAC clones for STS 9563 and STS 24967, both in the 3' utr of *PLAGL1/LOT1*. Full PAC clone names are given within the text. 1 and 2 are genomic control DNAs; B is a negative control PCR containing no template DNA. Both STSs are present in PAC clone 80-G20, but absent from all other PAC clones.

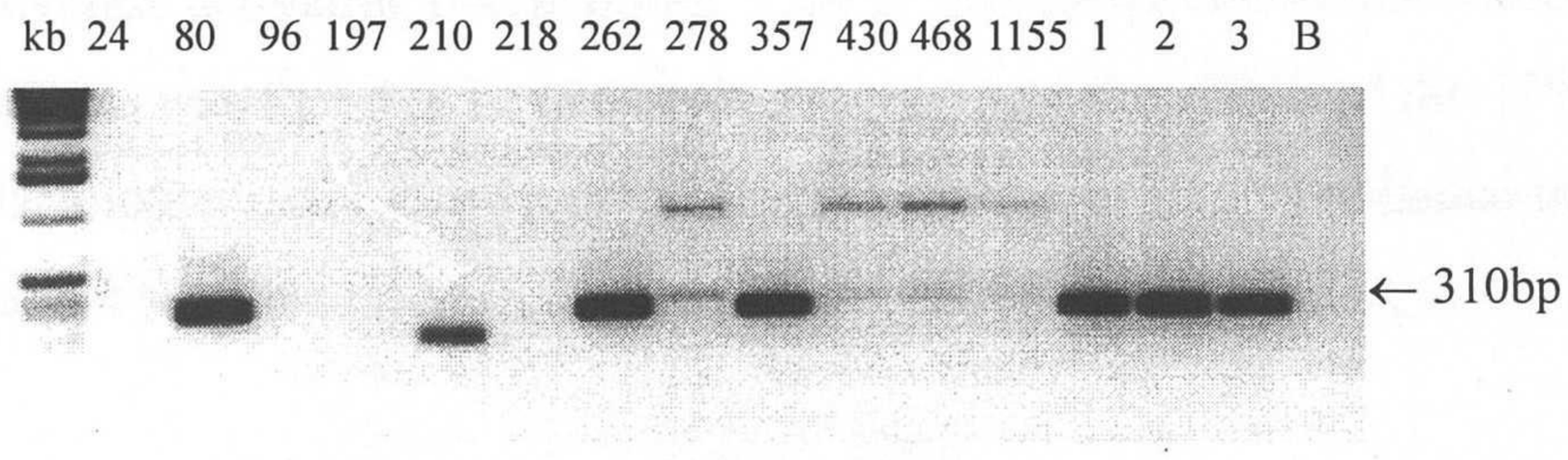


Figure 37. PCR screening of PAC clones for part of the coding region of *ZAC* (from 1927 – 2237, according to *ZAC* cDNA numbering, AJ006354). Full PAC clone names are given in the text. 1, 2 and 3 are genomic DNA controls; B is a negative control PCR containing no template DNA. PAC clones 80-G20, 262-B20, and 357-B1 are positive for this PCR.

4.2.11 *ZAC/LOT1/PLAGL1* – a summary of the literature describing its initial isolation and characterisation

In 1997, Spengler *et al*. (Spengler *et al.* 1997) cloned a novel mouse gene, encoding a zinc finger protein which regulates apoptosis and cell cycle arrest, which they consequently named *Zac1*. The protein was isolated by a functional expression cloning technique, which in this case identified clones that were able to induce expression of the gene encoding the type 1 receptor for the peptide PACAP (pituitary adenylate cyclase activating polypeptide). One of the cDNA clones that resulted from this screen was found to be identical in sequence to the cDNA of the *TP53* gene, whilst another represented a novel sequence, not previously published in any database, and it was this clone that represents the cDNA corresponding to the Zac1 protein.

Zac1 is a protein containing seven C_2H_2 zinc finger motifs in the N-terminal region. It is a member of a family of zinc finger proteins, although at the protein level it shows only 30% homology to its closest relatives, the GLI-Kruppel family. Other than the zinc finger motifs, Zac1 contains three other features of interest, which together are suggestive of a possible function as a transcription factor. These features are: a C-terminal region rich in amino acids proline, glutamine and glutamic acid, which is a common feature of transactivation domains of transcription factors; a putative phosphorylation site for cyclin dependent kinases between the second and third zinc finger motifs; and a putative protein kinase A phosphorylation site at the C-terminus, suggestive of regulation by protein kinases.

Spengler *et al*. demonstrated, by a colony-forming assay in soft agar, that Zac1 was capable of inhibiting proliferation of tumour cells *in vitro*, and that it could stop tumour formation in nude mice *in vivo*. In fact, their results showed that Zac1 was equally as effective as p53 in inhibiting tumour formation in xenografted nude mice. In short, the experiments of Spengler *et al*. demonstrated that Zac1 possesses an antiproliferative function, in its capacity to induce separately and independently the two processes of apoptosis and G1 arrest. Significantly, the gene encoding Zac1 is

the first identified gene, other than *TP53*, which can concurrently induce apoptosis and cell cycle arrest, but with separate control of these pathways. Of particular note is that Zac1 utilises a different mechanism than that of p53 (which is p21-dependent).

Independently, in 1997, Abdollahi *et al.* (Abdollahi *et al.* 1997a) cloned a novel rat gene, which they called *LOT1*. Their aim was to identify genes involved in ovarian cancer, and in order to do this they developed a rat tissue culture model of the disease. Their rationale was to use differential display to detect gene expression differences between normal cells, and the malignant transformants, both of which were derived from an identical genetic background. The technique successfully identified a transcript that was present in normal cells, but absent from the corresponding transformed cells, and therefore could feasibly represent a gene which was associated with ovarian tumorigenesis, since its expression was lost on transformation. The gene was accordingly named *LOT1* (lost on transformation 1). Abdollahi *et al.* cloned the human homologue of rat *LOT1*, and mapped it to the chromosomal region 6q25, which is a region known to be involved in the development of many tumours, including B-cell non-Hodgkin's lymphoma, ovarian cancer, breast cancer, melanoma, pancreatic cancer, and many others (Abdollahi *et al.* 1997b). In ovarian cancer, Abdollahi *et al.* themselves confirmed the allelic loss that has been reported at 6q25, and detected a 38% frequency of loss of heterozygosity for polymorphic markers in this region. Their experiments with human *LOT1* demonstrated that there are decreased or absent levels of expression of the transcript in both spontaneous and experimental human ovarian cancers, and they proposed *LOT1* as a tumour suppressor gene candidate.

In 1998, Varrault *et al.* (Varrault *et al.* 1998), who were the first group to identify the mouse gene *Zac1*, (Spengler *et al.* 1997), followed up their initial paper with a report of the isolation, chromosomal localisation, and functional characterisation of the equivalent human gene. They provided good evidence that the mouse and human genes were true orthologs, and found 74.6% identity between them at the nucleotide level, and 68.5% at the amino acid level. The main difference between the mouse *Zac1* and human *ZAC* cDNA sequences is the presence of a 34 proline-repeats

domain and a glutamic acid-cluster domain in the mouse sequence, both of which are absent from the human homologue. The presence of these repeats in the mouse gene but not in the human gene may offer an explanation as to the unusual phenomenon of a higher degree of nucleotide identity than amino acid identity between the mouse and human homologues. The degenerate nature of the genetic code predicts that the amino acid composition would be more similar between species than the nucleotide composition, but in this case, the calculated identity is perhaps distorted by the absence of the two large repeat regions which exist in the mouse gene.

It became apparent that human *ZAC* and human *LOT1* (Abdollahi *et al.* 1997b) were in fact the same gene, although they differed in their 5' untranslated regions. In addition, human *ZAC* is related to the *Plag* family of zinc finger genes, which are thought to have a possible role in tumorigenesis, particularly of pleiomorphic salivary gland adenomas. A new member of the *Plag* family of genes, *PLAGL1*, was cloned by Kas *et al.* (Kas *et al.* 1998), and like *LOT1*, this was found to be identical to *ZAC*, except for a different 5' non-coding region. Similarly, it was demonstrated that the C-terminal domain of *PLAGL1* can function as a positive regulator of transcription.

Varrault *et al.* (Varrault *et al.* 1998) performed *in vitro* experiments to establish the functional characteristics of *ZAC*. They discovered that, like the mouse homolog, human *ZAC* had transactivating and DNA-binding activity, and possessed antiproliferative properties, in its ability to inhibit tumour cell growth by induction of apoptosis and G1 arrest.

Varrault *et al.* (Varrault *et al.* 1998) gave the first defined map location of *ZAC* by FISH, and placed it between markers D6S308 and D6S978. Significantly, this is almost exactly the same region defined by Cave *et al.* (Cave *et al.* 2000), in their attempt to refine the putative TNDM locus at 6q24-25.

4.2.12 *ZAC* – a candidate gene for TNDM?

The location of *ZAC* in the critical TNDM region, and its role in cell cycle control make *ZAC* a very compelling candidate gene for TNDM. As already discussed, the gene responsible for TNDM is known to be a paternally-expressed imprinted gene, and it is the upregulation of this gene, either through paternal duplications of this region, or paternal uniparental disomy, that is believed to cause TNDM. Clearly, one of the main criteria for assessing candidate TNDM genes is to analyse the expression of the gene with respect to imprinting. It was decided therefore to determine whether *ZAC* is imprinted. A positive result would undoubtedly substantiate the candidacy of *ZAC* as the TNDM gene.

It was decided to test for imprinting of *ZAC* by determining whether *ZAC* shows monoallelic or biallelic expression. The rationale behind the experiment is to identify a sequence polymorphism in the coding region of *ZAC*, and to genotype a panel of first trimester fetal DNAs for this polymorphism. Details of this panel are given in Materials and Methods. Identification of a fetus that is heterozygous with respect to the polymorphism would allow the two parental alleles to be distinguished from each other. Furthermore, maternal DNAs corresponding to the fetal DNAs in the panel are available, and with knowledge of the maternal genotype, expression can be attributed to specific parental origin, should monoallelic expression be detected.

4.2.13 Search for sequence polymorphisms in coding region of *ZAC/PLAGL1/LOT1*

To identify a sequence polymorphism in *ZAC/LOT1/PLAGL1*, primers were designed in each of the exons, to amplify regions of the gene in several sections, to be sequenced radioactively. As explained above, there are three described cDNA sequences, all of which share some exons, but differ particularly in the 5' region upstream of the ATG start codon. A schematic diagram to explain the relationship between the three different transcripts is given in Figure 38.

PCR amplicons were designed to cover the different exons contained in the three different splice variants, concentrating particularly on the 5' untranslated region which, along with the 3' untranslated region, is more likely to contain a sequence polymorphism than the coding region. Figure 39 shows the location of the PCR primers used for polymorphism detection, and indicates the postitions at which single nucleotide polymorphisms were detected. Two polymorphisms were identified at residues 170 and 205 of the *LOT1* cDNA, and one at residue 875 of *ZAC*. The detection of the *ZAC* polymorphism at residue 875, through radioactive sequencing within the PCR product, is shown in Figure 40.

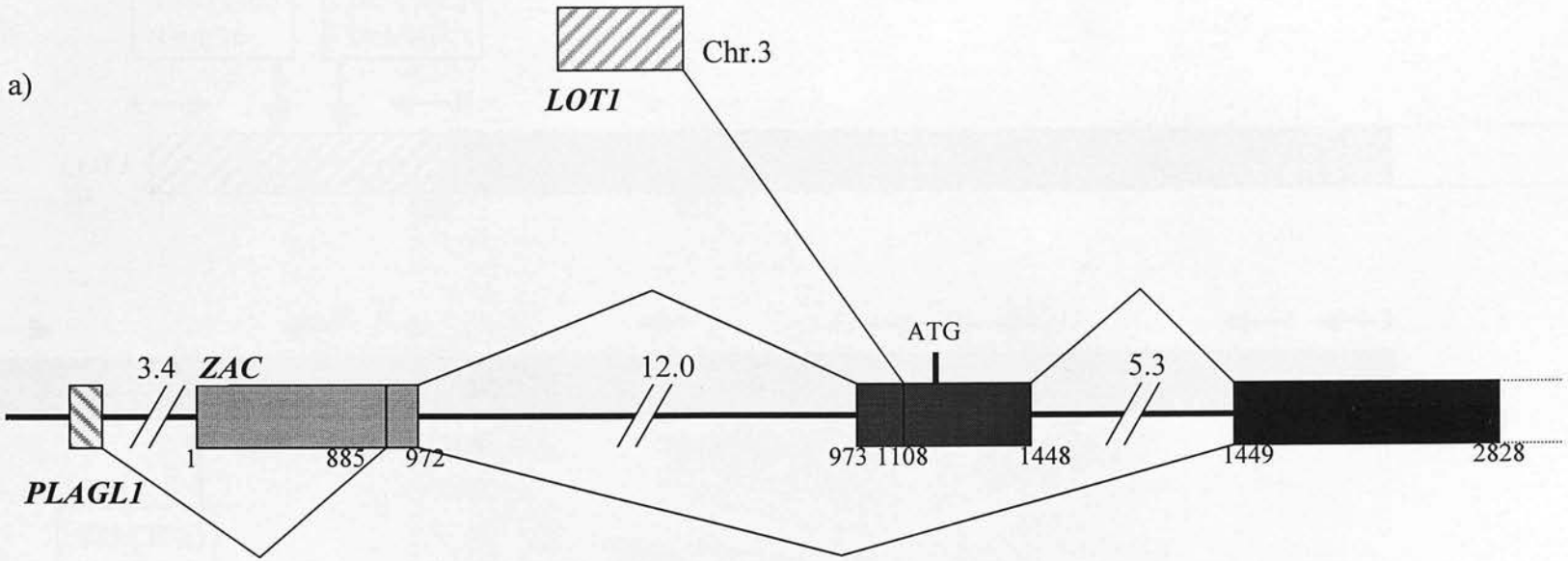


Figure 38. a) The splicing arrangement of the *ZAC*, *PLAGL1*, and *LOT1* transcripts. Intron sizes are given in kilobases. Numbering applies to the *ZAC* transcript. The first exon of *LOT1* was found to be derived from chromosome 3. b) The resulting transcripts. *PLAGL1* and *LOT1* have unique 5' exons, but share parts of the coding region of *ZAC*. The ATG is shown. Diagram not to scale.

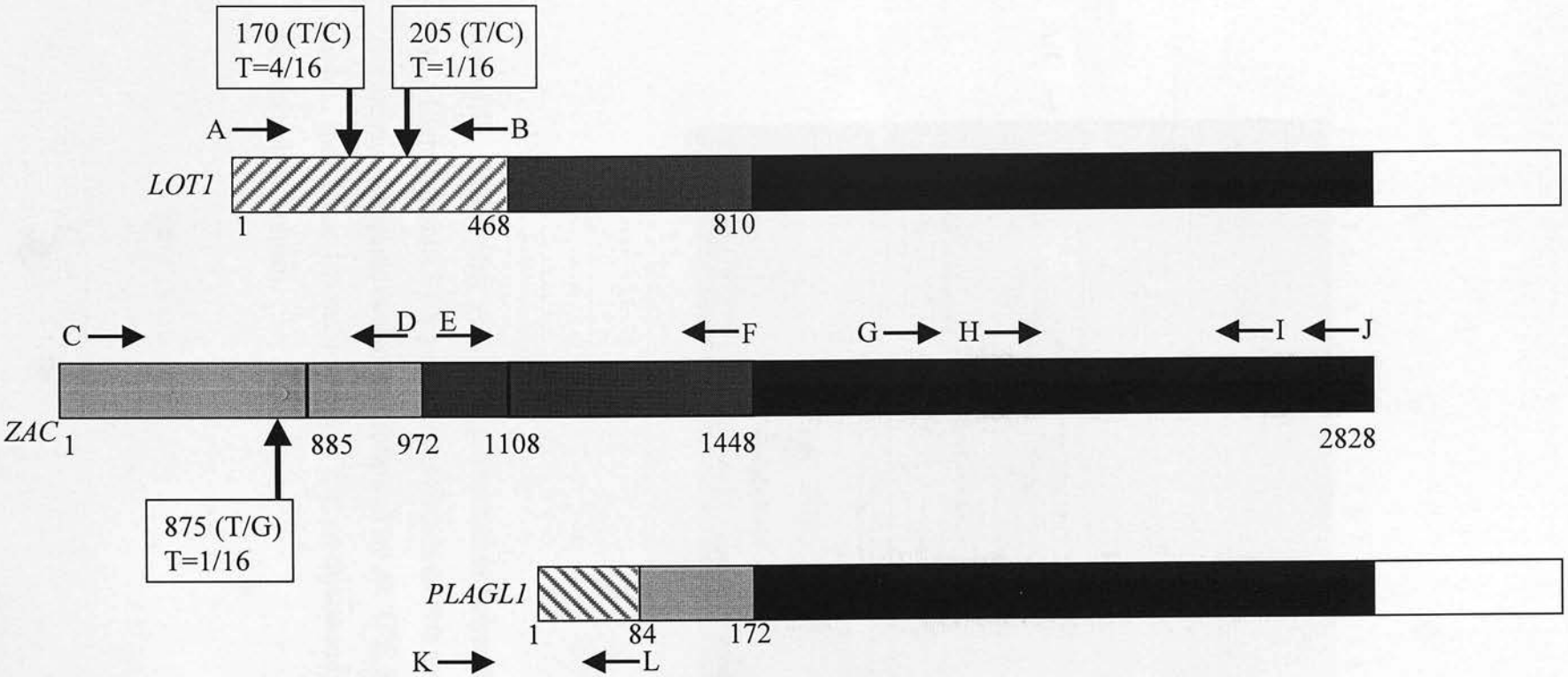


Figure 39. Detection of polymorphisms in the *LOT1* and *ZAC* sequences. The positions of the three polymorphisms are indicated, with details of their frequency and the nature of the substitution. Primer pairs used to amplify each region for sequencing analysis were A and B (LOT-A-5'/3'), C and D (ZAC-C-F1/ZAC-D-3'), E and F (ZAC-E-5'/ZAC-F-3'), G and I (hzacpoly1-5'/hzac2237-3'), H and J (hzac2018-5'/hzacpoly1-3'), K and L (PlagL1-5'/PlagL1-3'). Not to scale.

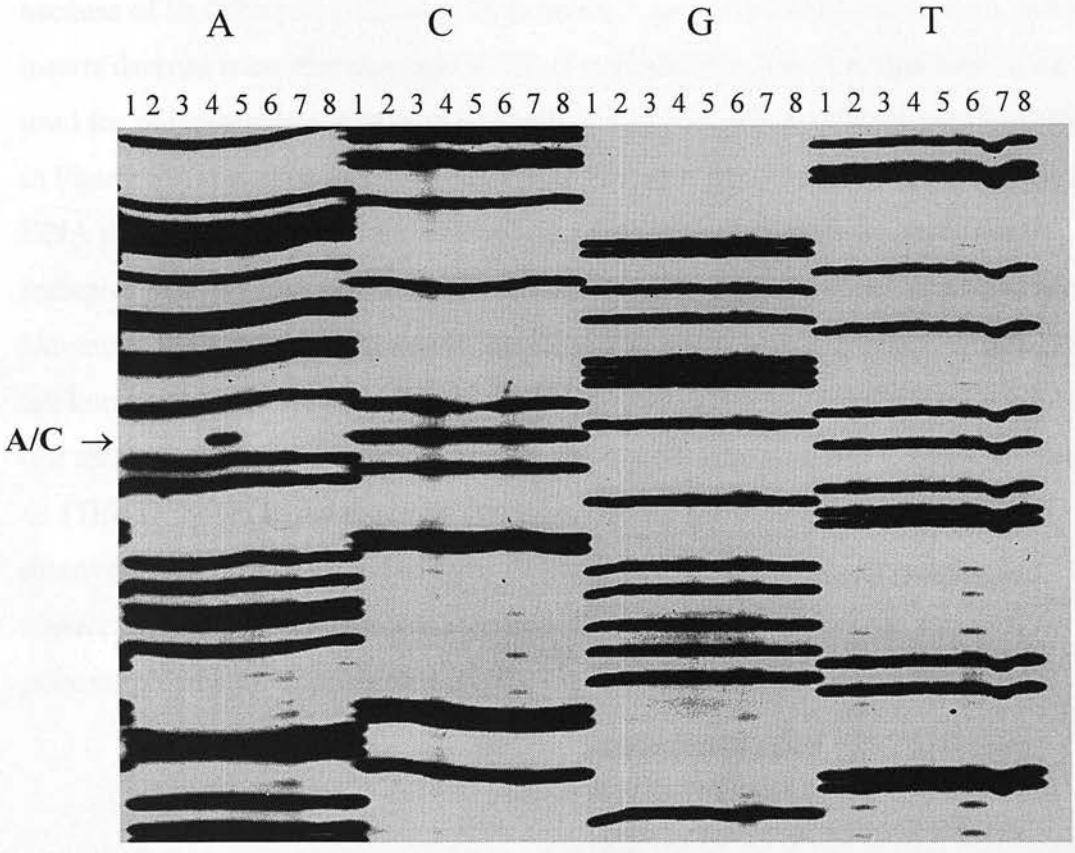


Figure 40. Sequencing gel showing detection of polymorphism at residue 875 of *ZAC*. Eight genomic PCR products are shown on this gel; for ease of detection, all 'A's are run alongside each other, followed by all 'C's, etc. Reverse sequence is shown. Individual 4 is heterozygous (A/C on this strand; T/G on the forward strand) for this polymorphism.

4.2.14 *LOT1* is a chimerism artefact

During the analysis described above, it became questionable whether the 5' utr of the published *LOT1* transcript, from residues 1 to 469, was actually on chromosome 6, because of its failure to generate a PCR product from PAC clones known to contain inserts derived from chromosome 6. To investigate this, the PCR that was being used for polymorphism detection (primers A and B (LOT-A-5'/3') located as shown in Figure 39) was used to screen the human monochromosomal Somatic Cell Hybrid DNA panel. The result of this PCR analysis is shown in Figure 41, and clearly indicates that this region is derived from chromosome 3, rather than chromosome 6. However, the subsequent exons of the *LOT1* transcript, which are shared with *ZAC*, are known to be located on chromosome 6, so it would seem that the *LOT1* transcript is a chimeric clone, comprising segments from both chromosome 3, and chromosome 6. (This is shown in the diagram of transcripts in Figure 38). Following this discovery, *LOT1* was excluded from further analysis, since it is not a genuine transcript. This unfortunately eliminated the possibility of using the two *LOT1* polymorphisms for expression analysis.

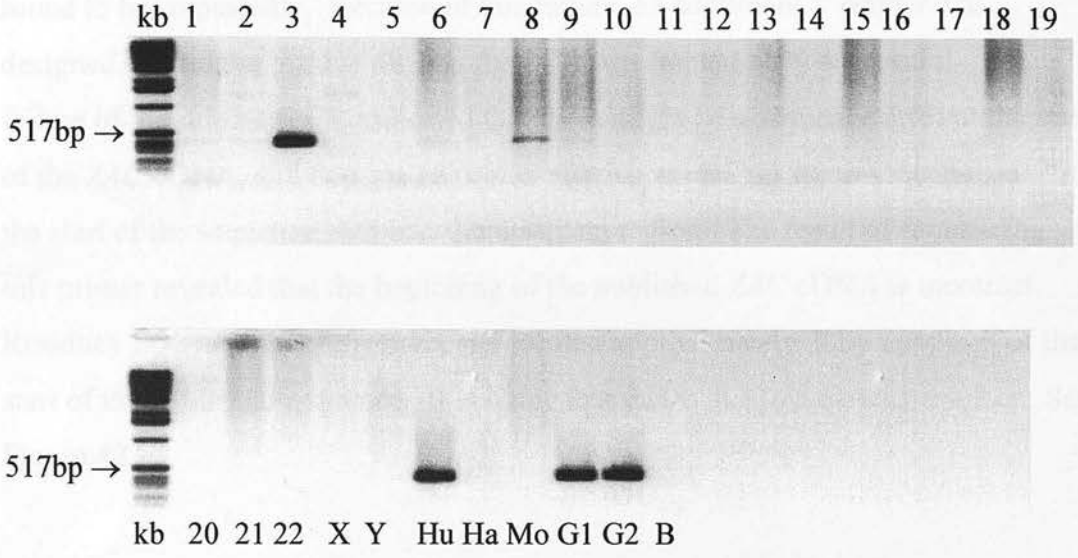


Figure 41. Screening of the somatic cell hybrid panel with a PCR at the 5' utr of *LOT1* (primers LOT1-A-5'/3'). This region of *LOT1* maps to chromosome 3. kb = 1 kilobase ladder marker, Hu = human genomic DNA control, Ha = hamster genomic DNA control, Mo = mouse genomic DNA control, G1 and G2 = further human genomic controls, B = negative control PCR.

4.2.15 The published *ZAC* cDNA sequence (AJ006354) contains an inversion

During the search for polymorphisms as described above, primers were designed to amplify several regions of *ZAC/LOT1/PLAGL1* from genomic DNA. One of these PCRs, with primers located at residues 3 and 965 of the *ZAC* cDNA sequence, was found to fail repeatedly. Because of this failure, an additional 5' primer was designed, starting at residue 68, and the PCR was immediately successful. The failure of the original PCR indicated that there might be a sequence error at the start of the *ZAC* cDNA, so I designed a primer starting at residue 89, to sequence towards the start of the sequence, and into the upstream region. The result of sequencing with this primer revealed that the beginning of the published *ZAC* cDNA is incorrect. Residues 1-30 are in fact inverted, and located approximately 30bp upstream of the start of the published sequence. It is likely that this is due to a cloning artefact. See Figure 42.

ZAC cDNA sequence, AJ006354:

```
1 cggcattttg ggacaactgt ttttaacgtt aataaatcac ttaggcgaga ataaattggc
  tttgttccat agcagatttg cctttgtact agttaagaaa atcctgaaaa gctttccct
```

Genomic sequence, PAC clone 340-H11, AL109755:

```
5'-tatatctcgcctaagtgatttattcacttgcctattttcttctgtgaatttgatcaaaaaggctatgtccatttattttg-
   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3'-atatagagcggattcactaaataagtgaacggataaaagaagacacttaaactagttttccgatacaggtaaataaaac-
                      ←34
                                                                4 →                33
-tatgagaacatcctttaagttaaaaatggaaaaaccaaggcattttgggacaactgttttttaacgttaatgcatcaca-3'
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
-atactcttgtaggaaaattcaatttttaccttttttggttccggtaaaaccctgttgacaaaaattgcaattacgtagtgt-5'
```

Figure 42. Evidence of a cloning artefact in the published cDNA sequence of *ZAC*. When compared to the genomic sequence, bases 4 to 33 of the cDNA are on the opposite strand, and in reverse orientation with respect to the remainder of the cDNA from base 34 onwards.

4.2.16 Expression analysis of *ZAC* by RT-PCR, using polymorphism 875

The G/T polymorphism identified at residue 875 of the *ZAC* cDNA sequence does not alter any restriction sites. However, a mismatched primer, ZAC-RT5', was designed starting at base 846, which creates a *Bsi*WI site when the polymorphic residue is G, but does not create this site when the polymorphic residue is T. For the purpose of genotyping a range of fetal DNAs for this polymorphism, a downstream intronic primer, ZAC-621-3' was used in conjunction with ZAC-RT5', to generate a PCR product of 200bp. Figure 43 shows the design of the genotyping PCR and RT-PCR, and the location of the primers with respect to the polymorphic residue 875. Digestion of the 200bp genotyping PCR product with *Bsi*WI yields bands of sizes171bp and 29bp in the case of a G/G homozygote, and bands of sizes 200bp, 171bp and 29bp in the case of a G/T heterozygote. DNA from 53 fetuses was analysed in this way; 4 were found to be G/T heterozygous (AC150452, HM170958, JC230671, and RT280475), and 49 were G/G homozygous. An example of the genotyping analysis is shown in Figure 44.

RT-PCR was performed on RNA samples available from tissues of the heterozygous fetuses using primers ZAC-RT5' and ZAC-RT3'(2), which generated a product of 290bp; RT-PCR did not generate a successful product in all cases, but it was not possible from this experiment to determine whether this was due to specific absence of *ZAC* expression in these tissues, or PCR failure due to the quality of the sample. The successful RT-PCR products that were obtained were from only three of the heterozygous fetuses, as detailed below. (RNA was available only from brain for Fetus RT280475, and since this was one of the failed samples, this fetus could not be included in the expression analysis).

Fetus AC150452: gut
heart
kidney
muscle

adrenal
cord

Fetus HM170958: muscle

Fetus JC230671: adrenal
lung
heart
kidney
muscle

Digestion of the RT-PCR products with *Bsi*WI gives bands of size 290bp if the polymorphic base 875 is T, and bands of sizes 261bp and 29bp if it is G. Biallelic expression from these heterozgous fetuses would be represented by bands of 290bp, 261bp and 29bp. The digestion products of the RT-PCR were separated on a 3% NuSieve agarose gel, shown in Figures 45 and 46. In all three cases, the RT-PCR products were of only one allele type with respect to the 875 polymorphism, despite originating from a heterozygous individual. This is clear evidence of monoallelic expression.

Unfortunately, due to the quality of the corresponding maternal DNA samples for the four heterozygous fetuses, attempts to determine the maternal genotype for the 875 polymorphism were unsuccessful, so it was not possible at this stage to establish whether the origin of *ZAC* expression is maternal or paternal.

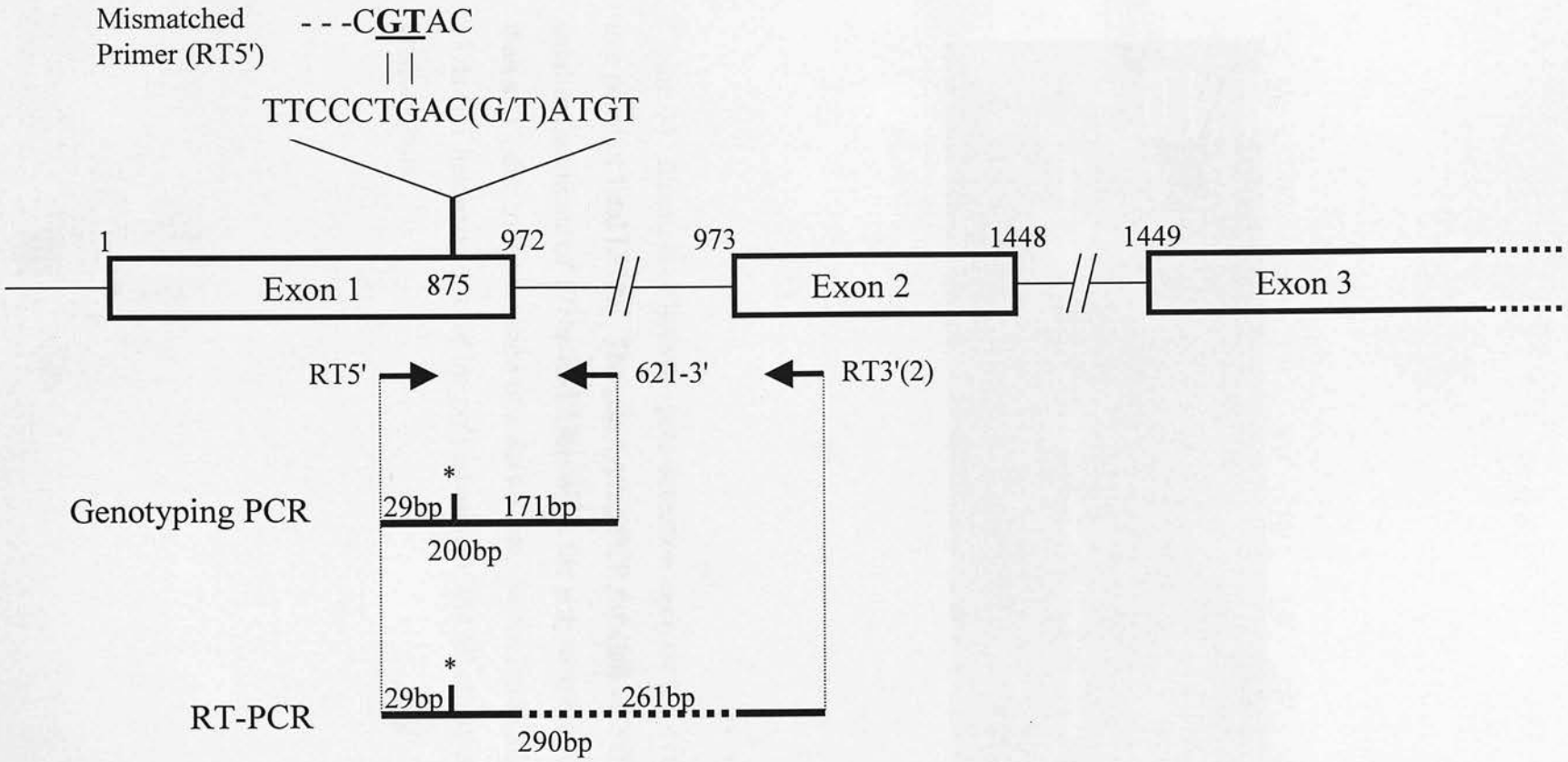


Figure 43. *ZAC* polymorphism 875 and the location of primers for the genotyping PCR and RT-PCR. Primer RT5' is a mismatched primer; the G and T residues of the primer shown in underlined bold are mismatches, in order to create a *Bsi*WI site (CGTACG) in the resulting PCR product when the polymorphic residue at 875 is a 'G' rather than a 'T'. The polymorphic residue is shown with an asterisk.

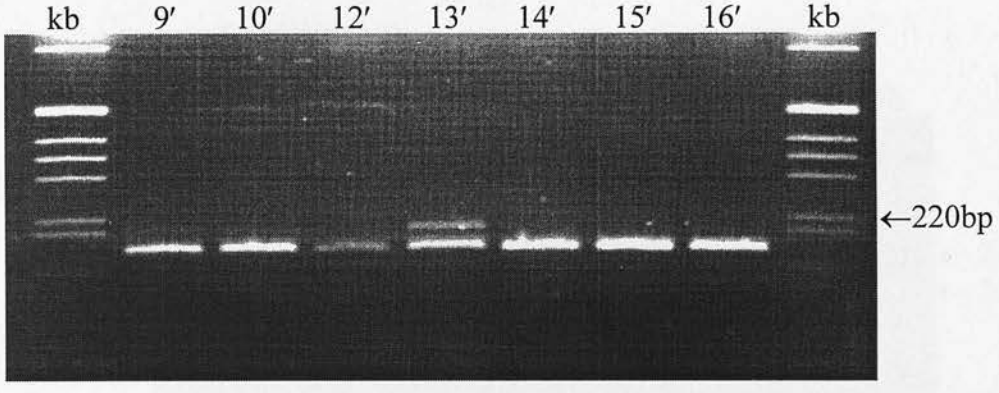


Figure 44. Example of heterozygote detection analysis for *ZAC* polymorphism 875 in a panel of fetal DNAs. The genotyping PCR product of 200bp is digested into two smaller fragments of 171bp and 29bp when the polymorphic residue is a 'G' rather than a 'T', due to the creation of a *Bsi*WI site. Seven fetal DNAs are shown; fetus 13' is G/T heterozygous for the polymorphism, and all others on this gel are G/G homozygous.

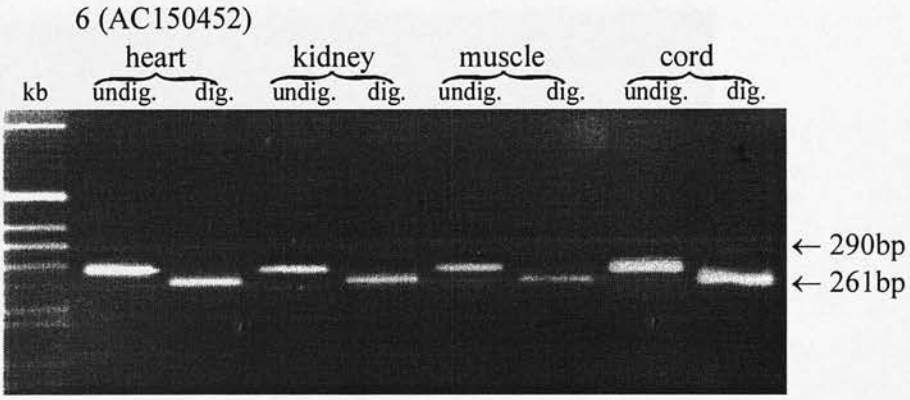


Figure 45. RT-PCR products for heterozygous fetus 6 (AC150452). In all tissues analysed, the RT-PCR product of 290bp is completely digested by *Bsi*WI into two fragments of 261bp and 29bp, suggesting monoallelic expression. Undig. = undigested, dig. = digested.

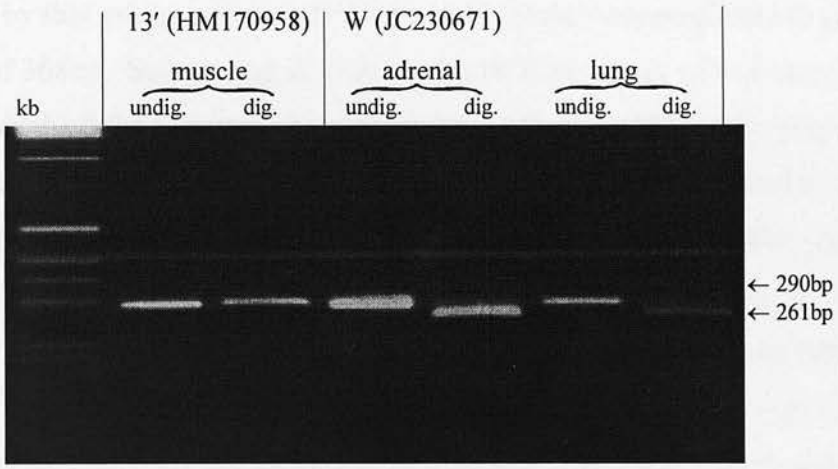


Figure 46. RT-PCR products for heterozygous fetuses 13' (HM170958) and W (JC230671). Undig. = undigested, dig. = digested. In the case of 13', all of the RT-PCR product remains undigested by *Bsi*WI, suggesting that it is derived solely from the 'T' allele. Conversely, in the case of W, there is complete digestion to the smaller fragments of 261bp and 29bp, suggesting that all of the product is derived from the 'G' allele which is cut by *Bsi*WI. Both results are consistent with monoallelic expression.

4.2.17 Expression analysis of *ZAC* by RT-PCR, using polymorphism 1029

A further polymorphism was identified in the laboratory of Y. Hayashizaki, at residue 1029 of *ZAC* cDNA. The nature of this polymorphism is a single nucleotide substitution of C for G. Using PCR primers PLAGL1ex2r2 and PLAGL1ex2f1 designed by that group, genomic DNA from 44 fetuses was amplified to give a product of 368bp. Sequencing directly on the PCR products, with primer ZAC-RT3'(2), enabled the fetuses to be typed for this polymorphism. The polymorphism is relatively common: 9 out of the 44 fetal DNAs analysed were found to be heterozygous at this locus. Maternal DNAs were analysed in a similar way, and two of the maternal DNAs, corresponding to heterozygous fetuses SH240167 and CM181169 were found to be homozygous (CC), making those fetuses fully informative. The maternal DNA for heterozygous fetus RT280475 was found to be heterozygous, rendering this sample uninformative in terms of tracing the parental origin of expression. For the remaining seven heterozygous fetuses, maternal DNA samples were either unavailable, or of inadequate quality and attempts to generate a PCR product failed consistently.

For the two informative fetuses, RT-PCR was performed on RNA from seven different tissues from each fetus. RT-PCR primers were ZAC1029-RT5' and ZAC1029-RT3', generating a product of 408bp. The design of the genotyping PCR and RT-PCR, along with the relative location of the primers is shown in Figure 47. The RT-PCR products were sequenced with primer ZACRT3'(2).

The following tissues were analysed:

Fetus SH240167: eye
kidney
heart
brain
gut

muscle
adrenal

Fetus CM181169: cord
gut
kidney
testes
brain
muscle
eye

In all cases, the RT-PCR product showed only the G allele when sequenced (see Figures 48 and 49), and since the maternal genotype was known to be CC, this indicates that expression of *ZAC* is solely from the paternal allele.

In the seven other heterozygous fetuses, RT-PCR was performed on a range of tissues. Since the maternal genotype is unknown, these are not informative as to the parental origin of the expressed allele, but can still be used to show monoallelic expression of *ZAC*. The following tissues were analysed by RT-PCR by the same method described above:

Fetus SS130773: adrenal
eye
kidney

Fetus JS020673: brain
stomach
kidney

Fetus SN301070: heart
muscle
cord

Fetus RT280475: muscle

Fetus HW150273: brain
heart
lung

Fetus CT111275: heart
muscle
gut

Fetus MS100972: muscle
liver
cord

In all cases, sequencing of the RT-PCR products showed only one allele type at the polymorphic locus 1029, and this is further evidence for monoallelic expression of *ZAC*. A summary of the evidence from these RT-PCR experiments for monoallelic expression of *ZAC* is given in Tables 7 and 8.

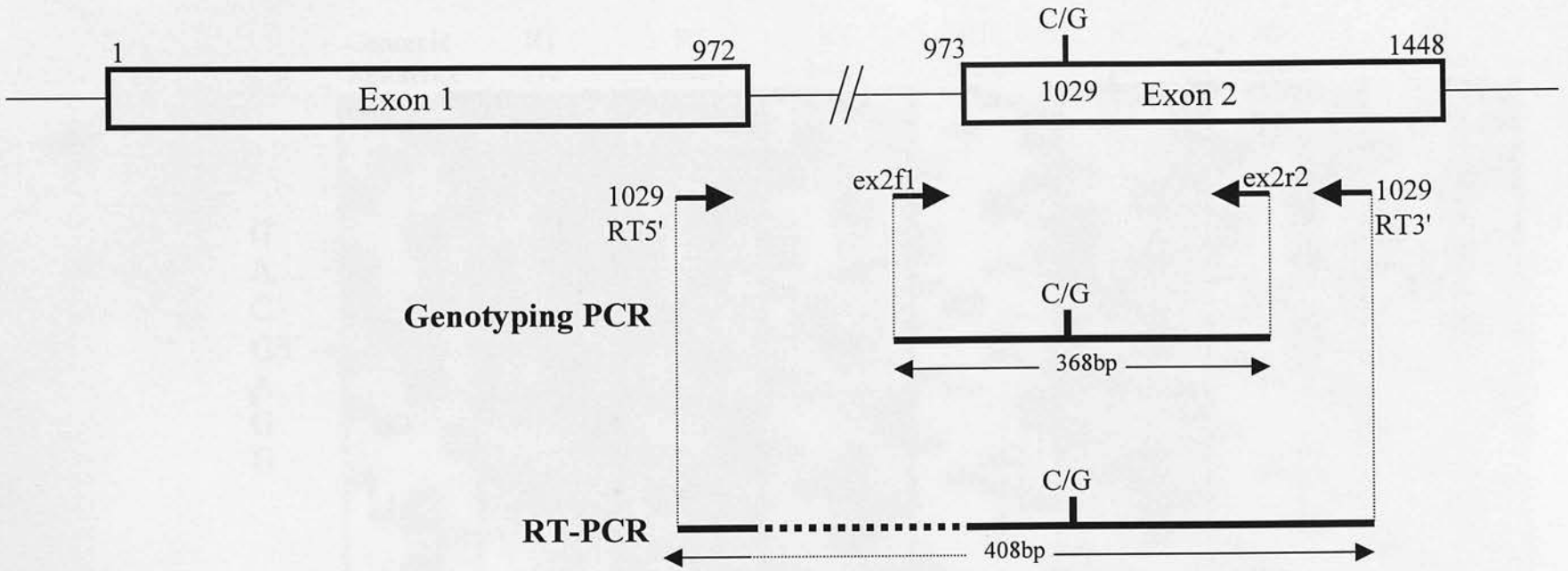


Figure 47. *ZAC* 1029 polymorphism, and location of primers for genotyping PCR (ex2f1 and ex2r2) and RT-PCR (1029RT5' and 1029RT3'). The polymorphic residue at 1029 is shown.

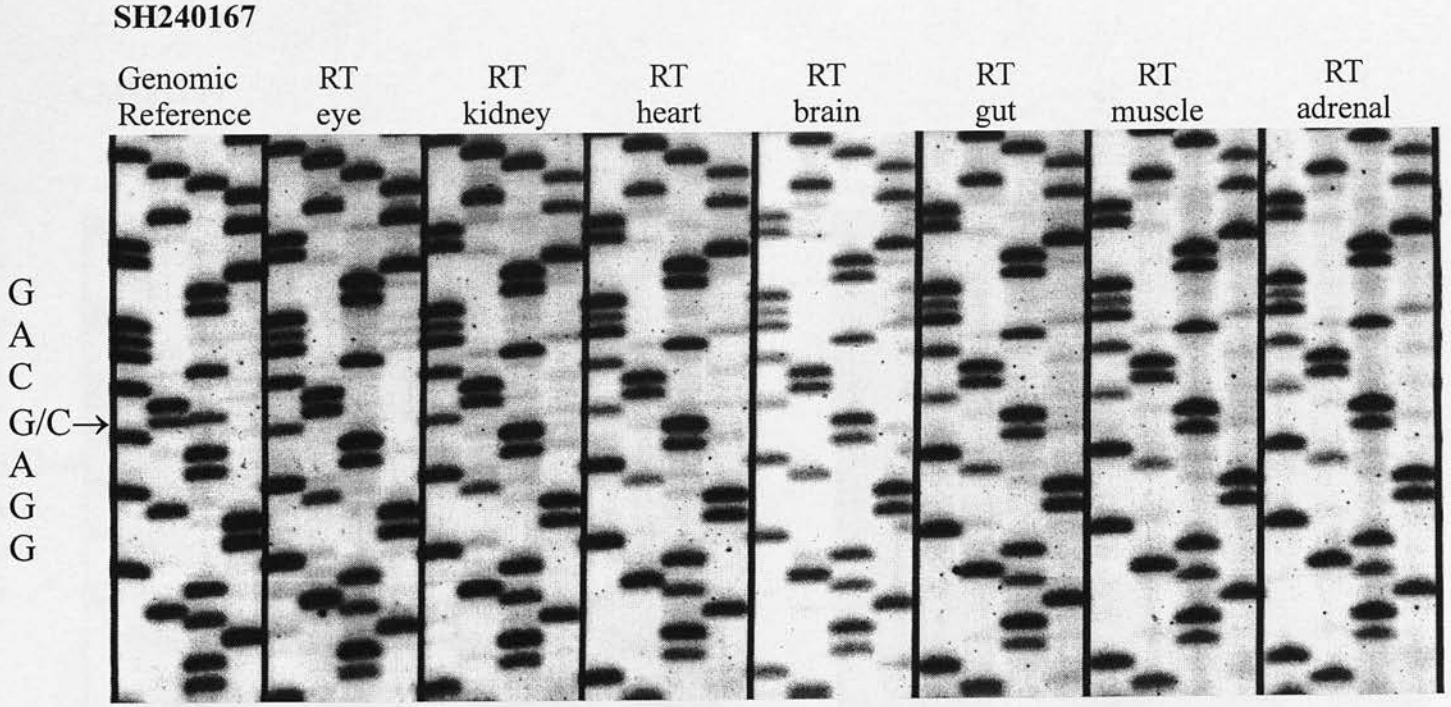


Figure 48. *ZAC* polymorphism 1029. Sequencing of a genomic reference PCR product of fetus SH240167 is shown in the first lane, to demonstrate heterozygosity, followed by sequencing of RT-PCR products from seven different tissues. In all RT-PCR products, only one allele is represented, indicating that expression is monoallelic. Sequencing is on the reverse strand using primer ZAC-RT3'(2).

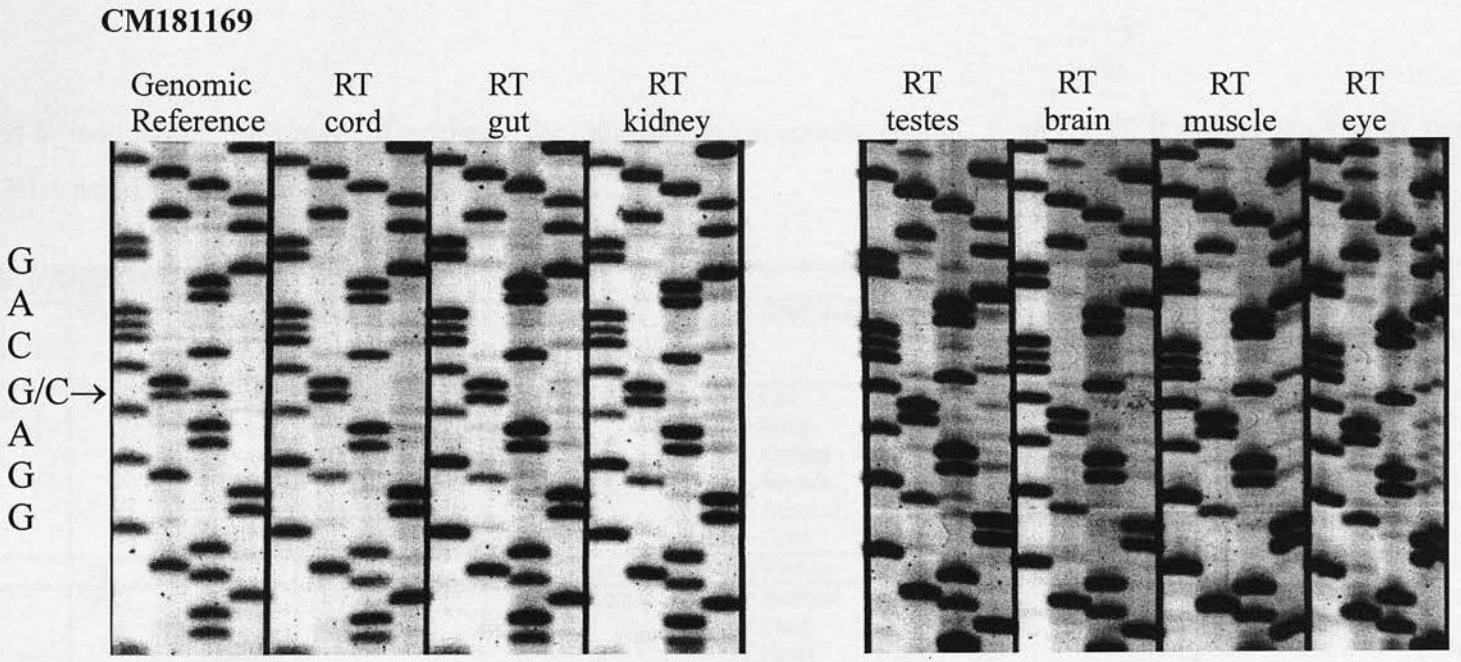


Figure 49. *ZAC* polymorphism 1029. Sequencing of a genomic reference PCR product of fetus CM181169 is shown in the first lane, to demonstrate heterozygosity, followed by sequencing of RT-PCR products from seven different tissues. In all RT-PCR products, only one allele is represented, indicating that expression is monoallelic. Sequencing is on the reverse strand using primer ZAC-RT3'(2).

Tables 7 and 8 (next page). Summary of evidence for monoallelic expression of *ZAC* from RT-PCR experiments using polymorphisms 875 and 1029. NI = non-informative.

ZAC POLYMORPHISM 875							
Fetus identification	**Fetus number**	**Genotype**	**Maternal Genotype**	**RNA analysed**	**Genotype of expressed allele(s)**	**Monoallelic expression?**	**Origin of expressed allele**
AC150452	6	G/T	Unknown	Gut	G	Yes	Unknown
				Heart	G	Yes	Unknown
				Kidney	G	Yes	Unknown
				Muscle	G	Yes	Unknown
				Adrenal	G	Yes	Unknown
				Cord	G	Yes	Unknown
HM170958	13'	G/T	Unknown	Muscle	T	Yes	Unknown
JC230671	W	G/T	Unknown	Adrenal	G	Yes	Unknown
				Lung	G	Yes	Unknown
				Heart	G	Yes	Unknown
				Kidney	G	Yes	Unknown
				Muscle	G	Yes	Unknown

ZAC POLYMORPHISM 1029							
Fetus identification	**Fetus number**	**Genotype**	**Maternal Genotype**	**RNA analysed**	**Genotype of expressed allele(s)**	**Monoallelic expression?**	**Origin of expressed allele**
SH240167	30'	C/G	C/C	Eye	G	Yes	Paternal
				Kidney	G	Yes	Paternal
				Heart	G	Yes	Paternal
				Brain	G	Yes	Paternal
				Gut	G	Yes	Paternal
				Muscle	G	Yes	Paternal
				Adrenal	G	Yes	Paternal
M181169	14	C/G	C/C	Cord	G	Yes	Paternal
				Gut	G	Yes	Paternal
				Kidney	G	Yes	Paternal
				Testes	G	Yes	Paternal
				Brain	G	Yes	Paternal
				Muscle	G	Yes	Paternal
				Eye	G	Yes	Paternal
SS130773	27'	C/G	Unknown	Adrenal	G	Yes	Unknown
				Eye	G	Yes	Unknown
				Kidney	G	Yes	Unknown
JS020673	28'	C/G	Unknown	Brain	C	Yes	Unknown
				Stomach	C	Yes	Unknown
				Kidney	C	Yes	Unknown
SN301070	A	C/G	Unknown	Heart	G	Yes	Unknown
				Muscle	G	Yes	Unknown
				Cord	G	Yes	Unknown
RT280475	H	C/G	C/G N.I.	Muscle	C	Yes	Unknown
HW150273	Q	C/G	Unknown	Brain	C	Yes	Unknown
				Heart	C	Yes	Unknown
				Lung	C	Yes	Unknown
CT111275	S	C/G	Unknown	Heart	C	Yes	Unknown
				Muscle	C	Yes	Unknown
				Gut	C	Yes	Unknown
MS100972	10	C/G	Unknown	Muscle	C	Yes	Unknown
				Liver	C	Yes	Unknown
				Cord	C	Yes	Unknown

4.2.18 *ZAC* is monoallelically expressed

The RT-PCR experiments detailed above using the polymorphisms at residues 875 and 1029 of *ZAC* show clearly that *ZAC* is monoallelically expressed in these fetal tissues, from only the paternal allele. However, these two polymorphisms are specific to the *ZAC* transcript. Neither of them are contained within the splice variant *PLAGL1*. Because of this, the RT-PCR experiments have only tested the origin of the *ZAC* transcript. In order to examine the expression status of *PLAGL1* with respect to imprinting, it would be necessary to identify an exonic polymorphism that was contained within the *PLAGL1* sequence, and to design primers specific to *PLAGL1*, so that the resulting RT-PCR products were solely of *PLAGL1* origin. Unfortunately, as detailed previously, an extensive polymorphism search was carried out, which identified only the *ZAC* polymorphism 875, and two *LOT1*-specific polymorphisms which are now known to be located in a chromosome 3-derived region of this chimeric transcript. *PLAGL1* contains a unique exon in the 5' utr, but no polymorphism was found either within this exon or in the 320bp upstream of this exon. The remainder of the *PLAGL1* transcript comprises common exons that are shared with *ZAC*, in which no polymorphism has been identified. See Figures 38 and 39.

It has therefore been impossible at this stage to determine the expression status of *PLAGL1*. The only option for rectifying this is to continue to search for exonic sequence polymorphisms. It would clearly be of great interest to investigate whether, like *ZAC*, *PLAGL1* is also subject to monoallelic expression.

4.3 DISCUSSION

4.3.1 Is *ZAC* the gene that causes TNDM when upregulated?

In this study I have performed mapping studies on the differentially methylated RLGS spot clone, NV149, and expression analyses of the nearby gene *ZAC*. NV149 contains a *Not*I site that is methylated exclusively on the maternally inherited allele, and is derived from a CpG island, which lies within 60kb of the recently identified gene *ZAC/PLAGL1/LOT1*. Given the chromosomal location of this gene, in a region known to contain a gene responsible for TNDM, it is logical to postulate that *ZAC/PLAGL1/LOT1* is a candidate for this role. From our knowledge of this disease, and the associated chromosomal abnormalities that have been detected in TNDM patients, there are four main criteria that must be satisfied by a genuine candidate. These are as follows:

1. The candidate gene must be imprinted, and expressed only from the paternal allele.
2. The candidate gene must reside in the smallest defined critical TNDM region between microsatellite markers D6S308 and D6S1010 (Cave *et al.* 2000).
3. The gene must encode a protein which possesses a function that could be envisaged to cause TNDM by a logical mechanism when upregulated.
4. TNDM patients must possess abnormalities of this gene such that they have an increased dosage of the gene product relative to unaffected individuals.

To address the first point above, I have performed RT-PCR experiments using two different polymorphisms in *ZAC*, and demonstrated monoallelic expression of the *ZAC* transcript in twelve first trimester fetuses, across a range of twelve different tissues. In two cases, the fetuses were fully informative with respect to the parental origin of each allele, and it was therefore possible to show in these cases that the monoallelic expression observed was derived from the paternal allele. These

experiments therefore provide strong evidence for *ZAC* being a paternally expressed imprinted gene.

Subsequent to these findings there were two further studies of relevance to the differential methylation of NV149 and the monoallelic expression of *ZAC*. Firstly, Gardner *et al.* (Gardner *et al.* 2000) mapped a 300-400kb candidate region containing the putative gene to the Sanger chromosome 6 contig 224, and then investigated the methylation status of the known CpG islands within this region. In PAC clone 340-H11, approximately 60kb upstream of the gene *ZAC*, they detected differential methylation of a CpG island partially overlapped by IMAGE clone 2073154. Their CpG island is in fact the same as the NV149 CpG island in this study, and thus confirms the differential methylation at this locus. Secondly, Piras *et al.* (Piras *et al.* 2000) identified *Zac1*, along with the epsilon sarcoglycan gene, *Sgce* through a subtractive screen for imprinted genes in the mouse. They demonstrated that both of these genes were maternally imprinted, and expressed solely from the paternal allele.

To summarise, there are now three independent studies which provide evidence either for differential methylation of the NV149/IMAGE 2073154 CpG island (Gardner *et al.* 2000), or for monoallelic expression of *Zac1* (Piras *et al.* 2000), or for both differential methylation of this CpG island and monoallelic expression of *ZAC* (Kamiya *et al.* 2000). Collectively, there is indisputable evidence that *ZAC* is imprinted and paternally expressed, thus fulfilling the first of the three criteria mentioned above. It is noteworthy, as an aside, that whilst *ZAC* is clearly imprinted and monoallelically expressed, it does not conform to the functional expectations of the parental conflict model of Moore and Haig (Moore, Haig 1991) (see Chapter 1), since it is paternally expressed, but is growth suppressing, rather than growth enhancing.

The second of the criteria that must be satisfied by candidate genes is the genomic location. Cave *et al.* (Cave *et al.* 2000) defined a small critical region at 6q24, flanked by polymorphic microsatellite markers D6S308 and D6S1010. The genomic

location of *ZAC* is within this interval, as evidenced by the location of the two markers STSG24967 and STSG9563, which are in the 3' utr of the gene, and lie between markers D6S308 and D6S1010 according to the Généthon map. Thus, *ZAC* satisfies the second of the criteria for candidate TNDM genes, and is currently the only known gene to reside in this region.

The third of the criteria is that a candidate TNDM gene must encode a protein which possesses a function that could be envisaged to cause TNDM by a logical mechanism when upregulated. As explained in detail in the introduction to this chapter, the genetic aberrations associated with TNDM are most commonly uniparental disomy for the paternal copy of chromosome 6, or paternally derived duplications of part of chromosome 6 including at least the 6q24 region. From these observed patterns of inheritance, it is apparent that TNDM is the result of upregulation of a paternally expressed imprinted gene. Because of this, it is expected that a candidate gene must play some role which could lead to TNDM if upregulated. *ZAC* encodes a protein that has been shown to induce both cell cycle arrest, and apoptosis, and one possible mechanism for its role in TNDM is the increased apoptosis of pancreatic β-cells that would result from upregulation of the gene, of which the main consequence would be significantly decreased insulin production. Another interesting aspect of *ZAC*'s function that may have pathological relevance is that it can induce expression of the gene for the type 1 receptor for the pituitary adenylate cyclase activating peptide (PACAP). This receptor is expressed in the pancreatic islets, and it is known that PACAP is produced by the pancreatic neural cells and is a highly active regulator of glucose-stimulated insulin release (Filipsson *et al.* 1999). The ability of *ZAC* to influence expression of the PACAP receptor may be important, because the consequence it may have on insulin release in the islet when overexpressed (as in paternal duplication or UPD6 cases) could relate directly to some of the pathology of TNDM. However, these ideas are theoretical, and remain to be investigated further.

It would seem that *ZAC* undoubtedly satisfies the first of the two criteria, since it is imprinted, paternally expressed, and lies within the critical TNDM interval, as defined by Cave *et al.* (Cave *et al.* 2000). In theory, it could satisfy the third criteria,

due to its function in inducing cell cycle arrest and apoptosis. As described above, this can be envisaged to result in TNDM when upregulated, since an increase in β-cell apoptosis would cause a severe reduction of insulin production leading potentially to TNDM. However, this idea is theoretical, and has not yet been investigated. It may be the ultimate cause of the permanent diabetes, which appears to be the eventual outcome in many cases of TNDM. Whether upregulation of *ZAC*, through modulating the PACAP-PACAP-R axis also directly mediates the earlier transient insulin deficiency seen in TNDM, is not clear at present. Clarification of this point will need a better understanding of the cellular physiology underlying PACAP-R signalling in the pancreatic beta cell.

The last of the criteria to be met by a candidate gene is that TNDM patients should display abnormalities of the gene, which result in an upregulation of its function. Such abnormalities could include paternal uniparental disomy and paternal duplications of all or part of chromosome 6, but the duplicated segment must specifically include the candidate gene. Patients lacking either of these abnormalities would therefore be expected to use alternative mechanisms by which the candidate gene is upregulated. These could include upregulating mutations in the sequence of enhancer or promoter elements, or epigenetic changes that reactivate the normally silent maternal allele. This fourth criterion is met partially by the fact that the duplications described to date include *ZAC/PLAGL1/LOT1*. However, as discussed below, paternal duplications and uniparental disomy do not account for all cases of TNDM.

According to Cave *et al.* (Cave *et al.* 2000), by pooling their data with that of Gardner's study, there is a total of 33 unrelated children with TNDM in the two studies. Of these, 27% have paternal uniparental disomy for chromosome 6, and 18% have a paternally derived duplication of this region. The remaining 55% of these affected children do not appear to have any detectable abnormality of this region. What is the basis of TNDM in these cases, and is *ZAC* or another gene involved?

Clearly, in order to show definitively that *ZAC* is the TNDM gene, there are experiments that must be done to address the points above. In particular, it could considerably strengthen the case for *ZAC* as a candidate TNDM gene if the unaccountable non-UPD and non-duplication cases of TNDM were found to contain either previously undetected submicroscopic duplications, or mutations, perhaps in the regulatory elements which would cause an upregulation of *ZAC*, or methylation mutations which would effectively give *ZAC* the paternal methylation pattern on both chromosomes and thus lead to the same result as paternal UPD6 or paternal duplications of the region. Gardner *et al.* (Gardner *et al.* 2000) addressed these questions in a recent study, and detected submicroscopic duplications in five patients in whom no apparent duplication had previously been detected. More significantly, they found that in two of thirteen patients with no chromosomal abnormality to account for TNDM there was an alteration of methylation, with both chromosomes showing a paternal, unmethylated pattern at the NV149 CpG island. This may represent the functional equivalent of paternal uniparental disomy, since both chromosomes carry a paternal epigenotype, and this would be predicted to cause an increased dosage of the gene product. The postulate that aberrant methylation could have a causative role in TNDM is substantiated by the lack of any similar defect in 50 normal control individuals (Gardner *et al.* 2000). It is noteworthy though, that the remaining eleven TNDM patients in Gardner's study are still unaccounted for by any of the proposed mechanisms for the disease.

ZAC is clearly a very strong candidate for the TNDM gene, but its role cannot yet be considered definitively proven. In particular, the issue of *ZAC*'s function in relation to TNDM should be investigated, and one obvious way to show directly that upregulation of *ZAC* causes TNDM would be to develop a transgenic mouse that overexpresses *ZAC*. However, the transient nature of the TNDM phenotype in humans, and the major differences in the developmental timeframe between the species, suggest that it might not be possible to reproduce the TNDM phenotype in a mouse model.

More generally, there are unanswered questions regarding the disease itself. How does TNDM predispose to adult diabetes? Why does the condition resolve itself during the first few months of life? Why are the symptoms variable between patients with the same genetic basis to TNDM (Marquis *et al.* 2000), even to the point that some cases go undiagnosed? Perhaps the answers to these questions will become more apparent once the mechanisms of *ZAC*'s involvement have been fully elucidated.

CHAPTER 5

IS *ZAC* A TUMOUR SUPPRESSOR GENE?

5.1 INTRODUCTION

5.1.1 The chromosomal region 6q is frequently deleted in cancer

Aberrations of the chromosomal region 6q have been associated with a broad spectrum of different tumour types in humans. In particular, deletions of 6q are found frequently in a variety of cancers, detectable by cytogenetic techniques or by microsatellite analysis for loss of heterozygosity (LOH). Early experimental evidence of a link between cancer and the complete or partial loss of chromosome 6q was revealed by the observation that this chromosomal region could reduce tumorigenicity in melanoma cells when introduced by microcell mediated chromosome transfer (Trent *et al.* 1989). Suppression of tumorigenicity by 6q was also observed in MDA-MB-231 breast cancer cells cells following similar chromosome transfer experiments (Negrini *et al.* 1994).

Evidence is mounting to suggest that the genetic basis to many cancers may be linked to the existence of several putative tumour suppressor genes at 6q, and the consequential loss of these genes by deletion and/or mutation, yet none of these putative genes has so far been definitively identified. Indeed, few candidates have been proposed to date. However, many of the studies have concentrated rather on a systematic approach of characterising the regions most commonly subject to deletion, and thus defining possible locations of these putative suppressor genes.

5.1.2 The involvement of 6q in breast cancer

Perhaps the most frequently studied cancer associated with loss of chromosome 6 is breast cancer. More than a decade ago, analysis of breast cancers by cytogenetic techniques and Southern blotting highlighted this chromosomal region as a hotspot (Mars, Saunders 1990; Dutrillaux *et al.* 1990), and subsequent studies were able to define distinct regions of LOH using polymorphic microsatellite markers. A study

by Noviello *et al.* (Noviello *et al.* 1996) utilised eleven polymorphic CA repeat markers to analyse 83 paired tumour and blood DNAs from breast tumour patients. The results of this analysis showed that the frequency of LOH was as high as 38.9% at some markers, and pointed towards three separate, common regions of allelic loss, defined by D6S251-D6S434 (at 6q14-21), D6S292-D6S311 (at 6q22-q23), and D6S441-D6S281 (at 6q24-q27). The idea of three commonly deleted regions at 6q was also suggested by the results of a study by Sheng *et al.* (Sheng *et al.* 1996), in which a 48% frequency of LOH was detected in a panel of 80 primary breast cancers. The commonly deleted regions in this case, however, were defined as D6S251-D6S252 (6q14-q16.2), D6S268-D6S261 (6q22), and D6S287-D6S270 (6q22). Whilst the first of the three regions defined by Sheng *et al.* is contained within the 6q14-q21 region defined by Noviello *et al.*, there is no overlap between any of the other regions identified in that study (Noviello *et al.* 1996). The relative locations of these markers and all others mentioned here are shown in Figure 50 which is a reference map derived from the LDB map at cedar.genetics.soton.ac.uk.

Another study, in which tissues from 66 cases of breast cancer were microdissected to eliminate any contaminating normal tissue, and then analysed with 15 polymorphic dinucleotide markers, detected allelic loss of either all or part of 6q in 80% of tumours (Fujii *et al.* 1996). This study defined a large common region of deletion at 6q23-q24, between markers D6S310/314 and D6S473/255, suggesting the presence of a tumour suppressor gene within this interval.

In the same year, Theile *et al.* (Theile *et al.* 1996) demonstrated suppression of tumorigenicity in the breast cancer cell line, CAL51, following microcell mediated transfer of chromosome 6. Subsequent analysis of the donor chromosome fragment in the resulting hybrids defined a critical region at 6q22-q23, containing the marker D6S310, and flanked by D6S292 and D6S311. This study also included an analysis of LOH for a panel of 24 polymorphic markers, in 46 primary breast cancer DNAs. The data suggested that there might be as many as four independent tumour suppressor genes mapping to 6q, one of which corresponds to the region defined by the chromosome transfer experiments.

As an alternative to microsatellite analysis, Zhang *et al.* (Zhang *et al.* 1998) investigated ten breast cancer cell lines using dual colour fluorescence in situ hybridisation (FISH). With this technique they were able to detect large regions of deletion extending from 6q12-q16 to 6q27, and a smaller deleted region in one cell line between 6q25.1 and 6q27, these results again suggestive of the notion that there may be more than one suppressor gene residing at 6q, though again not yielding results in close agreement with those of other studies.

Two of the more recent studies of the association between 6q loss and breast cancer are those of Utada *et al.* (Utada *et al.* 2000), and Rodriguez *et al.* (Rodriguez *et al.* 2000). Utada *et al.* examined a large sample of primary breast cancers for LOH at 16 microsatellite markers along 6q, and found allelic loss for at least one of these markers in 55% of the samples. Overall, their results defined two 1cM intervals, at D6S1040-D6S262 (6q22), and D6S305-D6S411 (6q25.3). More specifically, they were able to correlate losses at 6q22 with more aggressive tumour types (invasive solid tubular and schirrhous carcinomas), and correlate losses at 6q25.3 with loss of the progesterone receptor. They concluded therefore, that 6q22 and 6q25.3 represent the locations of two independent tumour suppressor genes involved in breast cancer, with particular relevance to tumour progression and loss of hormonal dependency respectively. A similarly large study was conducted by Rodriguez *et al.* (Rodriguez *et al.* 2000), who assessed the LOH in a panel of 178 paired breast tumour and normal DNAs using 30 polymorphic markers. The results of the study indicated a 76% frequency of allelic imbalance for at least one marker, and highlighted five distinct domains of allelic imbalance, defined by D6S300, D6S434, D6S261, D6S314-D6S409, and D6S441-D6S415.

From the many studies to date which have attempted to gain some understanding of the role of 6q in breast cancer, the overall conclusion is that this chromosomal region harbors at least one, and perhaps several, putative tumour suppressor genes, loss of which is strongly associated with tumorigenesis. There is considerable discordance, however, between the studies, both in the precise definition of the genomic intervals

believed to carry these genes, and in the determination of the true frequency of LOH at these locations. Indeed, whilst most studies have concluded that more than one gene is involved, individual studies have suggested that between one and five genes may play a role. As would be expected perhaps, those studies that have used a greater number of markers to detect LOH have been able to clarify the regions of deletion more precisely, and have consequently defined more separate, independent regions. For example, the studies of Theile *et al.* (Theile *et al.* 1996) and Rodriguez *et al.* (Rodriguez *et al.* 2000) employed 24 and 30 markers respectively, and identified 4 and 5 distinct regions of LOH, respectively. In contrast, studies such as those by Utada *et al.* (Utada *et al.* 2000) (16 markers) and Noviello *et al.* (Noviello *et al.* 1996) (11 markers) were less able to define accurately the true boundaries of the commonly deleted intervals, identifying only 2 and 3 distinct regions respectively. In addition to this issue, correlating the evidence from different studies is always hampered by the use of different genetic maps, between which exists a degree of discrepancy with respect to the true order of markers.

The wide variation in LOH frequencies found may also reflect the quality of the samples analysed. The problem of contamination of tumour samples with surrounding normal tissue is a source of variation between samples and between different studies, and may explain some of the discordance between the quoted LOH frequencies. For example, a higher LOH incidence than had previously been reported was detected in a study by Fujii *et al.* (Fujii *et al.* 1996), and was most likely due to their microdissection of tumour tissues to remove contaminating normal tissue, prior to microsatellite analysis.

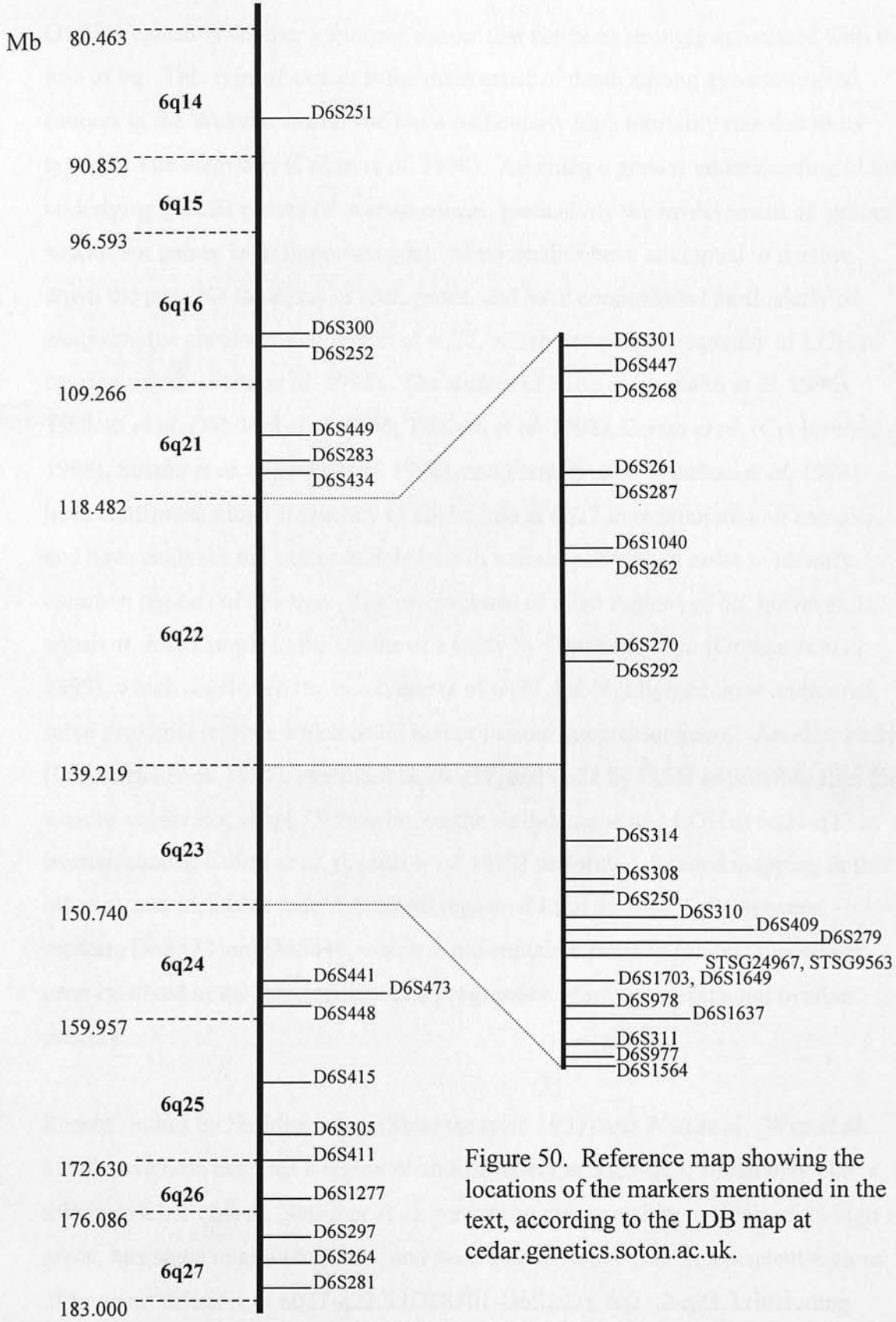


Figure 50. Reference map showing the locations of the markers mentioned in the text, according to the LDB map at cedar.genetics.soton.ac.uk.

5.1.3 The involvement of 6q in ovarian cancer

Ovarian cancer is another important cancer that has been strongly associated with the loss of 6q. This type of cancer is the main cause of death among gynaecological cancers in the Western world, and has a particularly high mortality rate due to its typically late diagnosis (Colitti *et al.* 1998). Attaining a greater understanding of the underlying genetic causes of ovarian cancer, particularly the involvement of tumour suppressor genes, is an important goal. Many studies have attempted to narrow down the possible locations of such genes, and have concentrated particularly on analysing the chromosomal region of 6q27, which has a high frequency of LOH in ovarian cancer (Saito *et al.* 1992). The studies of Saito *et al.* (Saito *et al.* 1996), Tibiletti *et al.* (Tibiletti *et al.* 1996; Tibiletti *et al.* 1998), Cooke *et al.* (Cooke *et al.* 1996), Suzuki *et al.* (Suzuki *et al.* 1998), and Foulkes *et al.* (Foulkes *et al.* 1993) have confirmed a high frequency of allelic loss at 6q27 in ovarian tumour samples, and have analysed the extent of deletions in tumour patients, in order to identify common regions of deletion. The involvement of other regions of 6q, however, is apparent, for example in the results of a study by Orphanos *et al.* (Orphanos *et al.* 1995), which confirmed the involvement of 6q27, but highlighted three additional, more proximal regions which could harbor tumour suppressor genes. Another study (Lastowska *et al.* 1994), identified 6q26-q27, and 6q24 by FISH as possible sites for tumour suppressor genes. Focussing on the well-documented LOH at 6q24-q27 in ovarian cancer, Colitti *et al.* (Colitti *et al.* 1998) performed detailed mapping in this interval, and identified a 4cM minimal region of LOH at distal 6q24 between markers D6S473 and D6S448, which could contain a putative tumour suppressor gene involved in the development and progression of invasive epithelial ovarian cancers.

Recent studies by Shridhar *et al.* (Shridhar *et al.* 1999) and Wan *et al.* (Wan *et al.* 1999) have both detected a region of similar extent at 6q23-q24, which may play a role in ovarian cancer. Shridhar *et al.* performed microsatellite analysis of 25 high grade, late stage ovarian tumours, and were able to define four independent regions of frequent deletion, at 6q21-q22.3 (D6S301-D6S292), 6q23.2-q23.3 (including

D6S978, D6S977, D6S311 and D6S1637), 6q26 (D6S411-D6S1277), and 6q27 (D6S297-D6S193). The second of these was a novel region spanning less than 500kb, containing the markers D6S311 and D6S977 which were found to have LOH frequencies of 89% and 71%, respectively. The study by Wan *et al.* involved microcell-mediated chromosome transfer experiments, to introduce a normal copy of chromosome 6 into the highly tumorigenic ovarian cancer cell lines HEY and SKOV-3. The resulting cells showed a significantly reduced tumorigenicity. Analysis of revertant subclones that had subsequently regained their tumorigenic potential enabled deleted regions to be identified, the smallest of which was defined by the marker D6S311, and flanked by markers D6S1649 and D6S1564, which were retained in the revertant subclones. The experiment thus identifies a region which is very likely to contain a tumour suppressor gene.

It is perhaps noteworthy that through many of these studies there is a recurrence of frequent LOH at the marker D6S311. Both Fujii *et al.* (Fujii *et al.* 1996) and Noviello *et al.* (Noviello *et al.* 1996) found D6S311 to be contained within their defined intervals of frequent LOH in breast cancer, and as mentioned above, the marker D6S311 defines minimal regions of deletion in ovarian cancer in at least two studies (Shridhar *et al.* 1999; Wan *et al.* 1999).

5.1.4 The involvement of 6q in lymphoma

The involvement of 6q in lymphoma was demonstrated in a cytogenetic analysis of 94 lymphoma patients, which, amongst several other common structural abnormalities, revealed that alterations of 6q were present in 31% of cases analysed. Even at this early stage in the investigation of 6q, the two distinct bands of 6q21 and 6q23 were found to be independently deleted in different cases of lymphoma (Bloomfield *et al.* 1983). Other studies have detected loss of 6q in B cell lymphomas at a frequency of 50% (Rickert *et al.* 1999), 38% (Monni *et al.* 1996), and 47% (Weber *et al.* 2000).

The chromosomal region 6q has also been implicated in T cell lymphomas (Schlegelberger *et al.* 1994; Knutsen 1998; Lepretre *et al.* 2000), and in acute lymphoblastic leukaemia (ALL), in which deletions of 6q have been narrowed down to the region at 6q14/15-6q21 (Menasce *et al.* 1994), and similarly, 6q16 (Merup *et al.* 1998), which is contained within the interval defined by Menasce *et al.*

In B cell lymphoma, the deletion of up to three independent regions of 6q has been detected, suggesting that more than one tumour suppressor gene may be involved in the pathogenesis of B cell lymphoma. A study by Gaidano *et al.* (Gaidano *et al.* 1992), identified two distinct regions of minimal deletion (RMD) at 6q25-q27, and 6q21-q23. The RFLP analysis used to define these intervals detected LOH in 22.5% of cases. A subsequent study by the same group further defined the regions of minimal deletion (RMD), and associated each with different pathological subsets of lymphoma, as follows: RMD1-6q25-27, RMD2-6q21, RMD3-6q23, associated with intermediate grade, high grade and low grade NHL respectively (Offit *et al.* 1993). A positional cloning approach was then taken by the same group in an attempt to map the location of the putative tumour suppressor gene residing in the RMD1 at 6q25-27 (Hauptschein *et al.* 1998). This generated a YAC contig across the whole region, and then defined a smaller minimal region of 5-9Mb believed to contain the gene. One candidate gene, *MLLT4* (also known as *AF-6*, and discussed further in 5.1.7), was found to reside in the RMD1, but is unlikely to be the true suppressor gene as no inactivating mutations were detected.

Several studies have now refined the frequently deleted regions in non-Hodgkin's lymphoma (NHL), and it is apparent that at least two, and possibly three tumour suppressor loci may be involved. The emerging consensus is that regions at 6q21, and 6q23-25 are most likely to contain tumour suppressor genes, along with a further possible locus at 6q26-27. The 6q21 interval has been identified in a comparative genomic hybridisation (CGH) study by Weber *et al.* (Weber *et al.* 2000) in a panel of primary diffuse large B cell lymphomas of the central nervous system, in which 47% LOH at 6q was detected. A 2 megabase interval of deletion flanked by markers D6S447 and D6S246 was detected in a FISH study using YAC probes spanning the

region at 6q16-q22 (Sherratt *et al.* 1997). A similar genetic location emerged from a study by Zhang *et al.* (Zhang *et al.* 2000), in which a commonly deleted region of 3cM (4-5Mb) was detected by FISH, and proposed to contain a tumour suppressor gene with a role in ALL, and both low and high grade NHL. A study by Starostik *et al.* (Starostik *et al.* 2000) detected 42% LOH of part or all of 6q in 31 large B cell lymphomas of gastric origin, using 73 polymorphic markers. In particular, this study defined two LOH hotspots, at 6q21-22.1 (flanked by D6S246 and D6S261), and 6q23.3-24 (flanked by D6S310 and D6S441), with 23% and 32% incidence of LOH respectively. As in other studies, the recurring 6q21 region was highlighted again, along with the less studied region at 6q23.3-24, which has also been detected as a site for a putative suppressor gene by Zhang *et al.* (Zhang *et al.* 1997). Interestingly, and in contrast to Offit *et al.* (Offit *et al.* 1993), the commonly deleted region at 6q23-q24 identified by Zhang *et al.* (Zhang *et al.* 1997) showed no discrimination between low or high grade and follicular subsets of lymphoma.

From the overall consensus of the results in these studies, it seems probable that there is a tumour suppressor gene at 6q21 involved in both NHL and ALL, and another tumour suppressor gene at 6q23-25 involved in NHL. The latter of the two corresponds well to the putative tumour suppressor loci proposed in other cancers, such as pancreatic, ovarian and breast cancer (see above and below), suggesting that within this interval may be a tumour suppressor gene with a common role in the tumorigenic pathway in a wide variety of malignancies.

5.1.5 The involvement of 6q in other cancers

Whilst the most frequently studied cancers in the context of 6q loss are breast, ovarian, and lymphoma, as discussed above, there are many others in which tumour suppressor genes at 6q may be involved. Some of these cancers are perhaps less significant in terms of their occurrence in the population, often reflected in the proportion of studies that have investigated them, but they are no less significant in terms of the frequency with which 6q is involved.

Pancreatic cancer has one of the poorest prognoses of all malignancies, so the elucidation of the genetic causes of this disease is a subject worthy of perhaps more study than it has received in recent years. A recent comparative genomic hybridisation (CGH) study showed 6q loss in 39% of endocrine pancreatic cancer cases (Speel *et al.* 1999). A large LOH study of 55 paired pancreatic tumour DNA and corresponding normal DNA samples with 30 polymorphic markers revealed that three independent regions of 6q might contain tumour suppressor genes (Abe *et al.* 1999). These comprise firstly a 500kb region at 6q21 bordered by D6S449 and D6S283, showing a LOH frequency of 69%; secondly, a 7cM region at 6q22-23 bordered by D6S292 and D6S308, showing 60% LOH; and thirdly, a 13cM region at 6q25-q27 bordered by D6S305 and D6S264, showing 51% LOH. It is possible that all three regions may represent tumour suppressor gene loci.

In cervical cancer, loss of 6q has been reported at a frequency of 38% (Kirchhoff *et al.* 1999), and a study by Mazurenko *et al.* (Mazurenko *et al.* 1999) predicts the existence of possibly three tumour suppressor genes, located at 6q23 (D6S311), 6q26, and 6q27, with LOH frequencies of between 30 and 40% in each case.

Papillary serous carcinoma of the peritoneum was recently found to be strongly associated with the loss of 6q. As with cervical cancer, three minimal regions of LOH have been detected and defined as 6q23-q24 (D6S250-*ESR*, containing D6S311), 6q25.1-q25.2, and 6q27 (Huang *et al.* 2000). The three intervals show LOH frequencies of 42.9%, 47.2% and 47.4% respectively.

Cytogenetic analyses established the involvement of 6q losses in the *in vitro* transformation of uroepithelial cells (Wu *et al.* 1991). Subsequent to this, a comparative genomic hybridisation (CGH) study demonstrated that 6q losses were among the most frequent of genetic alterations in bladder cancer, occurring in 19% of cases analysed (Richter *et al.* 1998). Similarly, LOH at 6q was shown to occur at a frequency of 27% by Bernues *et al.* (Bernues *et al.* 1999). In particular, loss of the 6q22-q23 region has been associated with the progression of bladder cancer, suggesting that a tumour suppressor gene resides in this interval (Richter *et al.* 1999).

Salivary gland carcinoma has shown a 47% LOH frequency in a study by Queimado *et al.* (Queimado *et al.* 1998), and two small, commonly deleted regions have been defined at 6q21-q23.3, and 6q27.

Three common regions of allelic loss have been detected in malignant mesothelioma. These are defined by 6q14-q21, 6q21-q23.2, and 6q25 (Bell *et al.* 1997).

Allelic loss of 6q has been found in prostate cancer at frequencies of 22% (Visakorpi *et al.* 1995), 39% (Cher *et al.* 1996), and 33% (Cooney *et al.* 1996). Similar regions of common deletion were detected by (Cooney *et al.* 1996) and (Srikantan *et al.* 1999), representing the interval from 6q14/16.3 to 6q21. Additionally, Srikantan *et al.* detected allelic loss at D6S311 (6q23), which may represent an additional locus with some role in prostate cancer. Interestingly, this is a yet another example of the recurrence of the marker D6S311; this marker appears to signify a site of frequent LOH in many cancers, including non-Hodgkin's lymphoma (Starostik *et al.* 2000), papillary serous carcinoma of the peritoneum (Huang *et al.* 2000), and breast and ovarian cancers as mentioned previously, in addition to prostate cancer.

5.1.6 Summary

There is clearly a general consensus that 6q harbors several tumour suppressor genes, but correlating the data from many different studies is difficult. Overall, it would appear that there may be up to four putative tumour suppressor loci at 6q16, 6q21, 6q23-q25, and 6q27, but whether the loci are specific to certain cancer types, or disease stages, is unclear. However, the fact that similar deleted regions recur in many studies, across many cancers, suggests that one or more of these tumour suppressor genes may have a common role in tumorigenesis.

5.1.7 Candidate tumour suppressor genes at 6q

Despite the accumulating evidence to suggest that several tumour suppressor genes reside on the long arm of chromosome 6, few candidates have yet been proposed. One candidate is the gene *SOD2*, which is located at 6q25.3 and encodes the protein manganese superoxide dismutase. This gene has been proposed as a candidate tumour suppressor gene in melanoma (Bravard *et al.* 1992). However, the evidence for a suppressor role is conflicting, and there is experimental evidence against the ability of this gene to suppress tumorigenicity from transfection studies that showed that the introduction of a normal chromosome 6 (and intact copy of *SOD2*) into metastatic melanoma cell lines did not suppress tumorigenicity or metastatic potential (Miele *et al.* 1995).

Another candidate is the gene *AF-6/MLLT4*, which was found to reside in the RMD1 described in lymphoma at 6q27 (Hauptschein *et al.* 1998). Many acute leukemias and acute lymphoblastic leukemias are found to have translocations that break at the 11q23 region, and result in fusion with another chromosomal region. The gene *AF6* is one of four common fusion sites involved in this translocation, hence its name, *AF6* (ALL1 Fused Gene from chromosome 6), or *MLLT4* (Myeloid/Lymphoid or Mixed Lineage Leukemia, Translocated to, 4) (Prasad *et al.* 1993). Despite its apparent association with leukemia, no muations were identified in a small subset of B cell NHL cases (Hauptschein *et al.* 1998). This gene also lies within a commonly deleted region defined by Saito *et al.* (Saito *et al.* 1996), but similarly, in that instance, no mutations of the gene were detected in ovarian cancer, thus suggesting it unlikely that *AF-6/MLLT4* is a tumour suppressor gene in these cancers.

Other candidates are the oestrogen receptor *ESR* at 6q25, the presence of which has been established as an identifying marker for breast cancer patients who have lower risk of relapse and a generally good prognosis (Clark, McGuire 1988); and *MAS1*, which encodes an integral membrane protein that may function in cell signalling. *MAS1* is an oncogene originally identified from an epidermoid carcinoma cell line (Young *et al.* 1986), and maps to tumour-specific breakpoints at 6q24-27 (Rabin *et al.* 1987). It had been shown to be imprinted by Villar *et al.* (Villar, Pedersen 1994),

but this was later revealed to be incorrect, as the analysis of imprinting by Villar *et al.* was likely to have been distorted by the fact that the last exon overlaps an imprinted antisense transcript of the *Igf2r* gene (Lyle *et al.* 2000). No compelling evidence has come to light as yet to substantiate the involvement of either *MAS1* or *MLLT4* in a tumour suppressor capacity. The identification of the gene *ZAC* at 6q24, and the accompanying evidence of its anti-proliferative properties was therefore an exciting discovery in the search for candidate genes in this region.

5.1.8 *ZAC* as a tumour suppressor gene in breast cancer

The gene *ZAC*, also known as *PLAGL1* and *LOT1*, has been proposed as a candidate tumour suppressor gene, for three reasons. Firstly, *ZAC* encodes a zinc finger protein, which has DNA-binding, transactivation, and antiproliferative properties, and possesses the ability to concurrently induce apoptosis and G1 arrest (Spengler *et al.* 1997), which is reminiscent of the action of p53, although the relevant pathways are believed to be distinct for the two genes. Secondly, *ZAC* was identified independently as *LOT1*, through a screen for transcripts specifically lost on transformation of normal ovarian cells to cancer cells (Abdollahi *et al.* 1997a), suggesting that it acts as a suppressor of tumorigenicity in ovarian cells. Thirdly, *ZAC* resides in a region of 6q23-25 that has been implicated in many human cancers, as described in 5.1.2-5.1.5. Specifically, the *ZAC* locus falls within the critical regions defined by (Noviello *et al.* 1996) (breast), (Theile *et al.* 1996) (breast), (Starostik *et al.* 2000) (large B cell NHL), and (Huang *et al.* 2000) (papillary serous carcinoma of the peritoneum), and is in very close proximity to critical regions defined by (Abe *et al.* 1999) (pancreatic), (Shridhar *et al.* 1999) (ovarian), (Wan *et al.* 1999) (ovarian), and (Rodriguez *et al.* 2000) (breast). Figure 51 shows the location of *ZAC* relative to these defined critical regions, which are indicative of a tumour suppressor gene locus. No such gene has been identified in this region to date, and for the reasons stated above, *ZAC* is a strong candidate tumour suppressor gene.

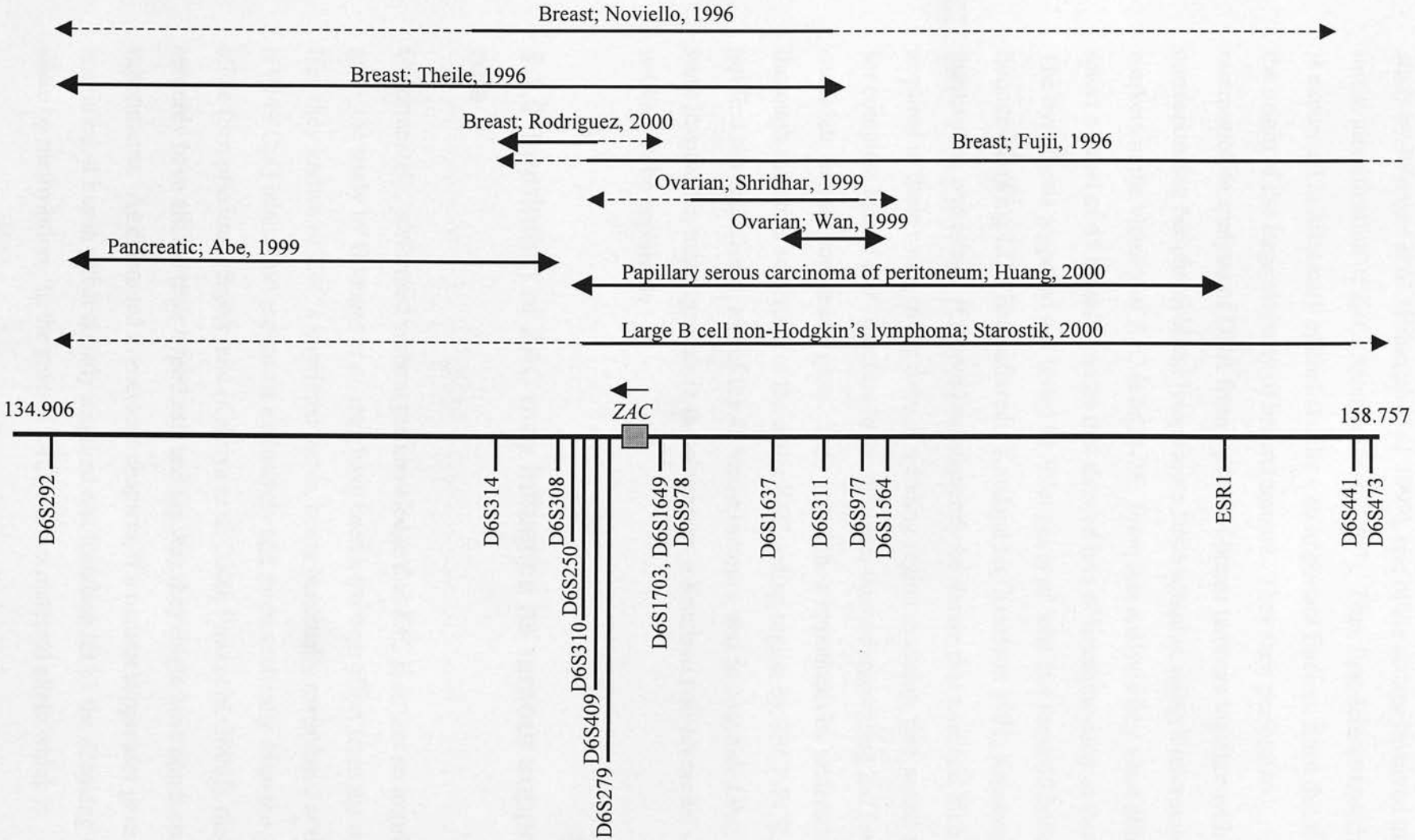


Figure 51. A schematic diagram to illustrate some of the many defined regions of deletion at 6q22-24 in different cancers. Order of markers is according to Cedar LDB map. See text for details of each study. Solid lines indicate the intervals defined by the LOH of markers included in the study; dotted lines indicate the potential extent of each interval given that not all markers were analysed, and some were uninformative. Scale is in Mb.

The candidacy of *ZAC* as a tumour suppressor gene in breast was addressed in a large study by Bilanges *et al.* (Bilanges *et al.* 1999), one of the groups involved in the initial identification of *ZAC* (Spengler *et al.* 1997). They first demonstrated that *ZAC* is expressed in mammary epithelial cells – an important finding, since this tissue is the origin of the large majority of breast tumours. They then performed microsatellite analysis of DNA from a panel of breast tumours together with their corresponding peripheral blood lymphocyte DNA samples, using 5 microsatellite markers in the vicinity of *ZAC* at 6q24-25. From this analysis they were able to select a panel of 45 breast tumours that showed loss of heterozygosity at this locus. The hypothesis proposed and tested by Bilanges *et al.* was that these 45 breast tumours showing LOH had suffered a Knudson hit (Knudson 1971; Knudson 1986), inactivating one allele. If it could subsequently be shown that a second hit had been acquired in these cases, in the form of a coding region mutation, this would account for complete loss of *ZAC* function in the tumours, thus substantiating *ZAC* as a candidate tumour suppressor gene. They tested their hypothesis by undertaking thorough mutation screening of the entire *ZAC* coding region by SSCP-PCR, but did not find any mutation in any of the 45 breast tumours, and so concluded that if *ZAC* were involved in tumorigenesis in these samples, a Knudson two-hit model was unlikely to be applicable.

5.1.9 Imprinting of *ZAC* may influence its tumour suppressor role

Unfortunately, performed without the knowledge that *ZAC* is in fact an imprinted gene, the study by Bilanges *et al.* may have been a fruitless effort from the outset. Had they known of *ZAC*'s imprinted status, being maternally methylated at the NV149 CpG island and expressed exclusively and monoallelically from the paternal allele (See previous chapter, and (Kamiya *et al.* 2000; Piras *et al.* 2000)), this would certainly have altered their hypothesis and the way they might have conducted their experiments. As discussed in previous chapters, if a tumour suppressor gene is imprinted, it has in effect already acquired one Knudson hit by the silencing of one allele by methylation. In the case of *ZAC*, it is the maternal allele which is

consistently methylated, therefore leaving only the paternal allele in an active state. This in itself is a precarious situation for a tumour suppressor gene, since the subsequent loss of the one remaining allele is all that is required for a complete loss of function, which could lead to tumorigenesis. In the panel of breast tumours selected in the Bilanges study showing LOH at 6q24-25, it is perhaps not surprising that no mutation was detected in the remaining allele, because it is quite feasible that the lost allele was the active paternal copy, and the retained allele was the inactive maternal copy in all cases.

5.1.10 Purpose of this study

With the knowledge that *ZAC* is imprinted, a new hypothesis can be tested. In tumours with LOH, if it can be shown that the single retained allele is consistently of maternal origin and thus inactive, then this would be strongly suggestive of a tumour suppressor role for *ZAC*.

In a previous chapter within this study, I have demonstrated the monoallelic expression and imprinting of *ZAC* by RT-PCR. Confirmation of this finding has recently been published in a study by another group who independently identified *Zac* along with another maternally methylated gene in a systematic screen for imprinted genes in the mouse (Piras *et al.* 2000). With this substantiated knowledge of *ZAC*'s expression status, an important hypothesis to test now would be whether the origin of the retained allele is consistently maternal in a panel of tumours with LOH at the *ZAC* locus.

To test this hypothesis it would first be necessary to analyse a panel of tumours using a range of microsatellite markers close to the *ZAC* gene. On establishing a subset of tumours showing LOH, the origin of the remaining allele would be determined by analysing the methylation of the NV149 CpG island upstream of *ZAC*, using the technique of bisulphite modification and sequencing. The expected result is that tumours with LOH at this chromosomal region would show only the maternal methylation pattern, indicating that the single allele that they have retained is

methylated and thus functionally silent. A particularly important aspect of the experiment is that the methylation status of the subset of tumours without LOH should be tested; to remain consistent with the proposed hypothesis, these tumours should display both a maternal and paternal methylation pattern at the CpG island, since they possess both parental alleles. This point would show that the exclusively maternal methylation pattern observed in LOH tumours is specifically due to the LOH of the paternal allele, rather than any silencing of the paternal allele through tumour-specific methylation. A recent RLGS study has assessed the frequency of aberrant methylation of CpG islands in different tumour types and concluded that on average, 600 of the estimated total of 45000 CpG islands in the human genome were aberrantly methylated in tumours. However, the results of this study also indicated that aberrant methylation was rarer in certain types of tumour, including breast, head and neck, and testicular tumours, whereas it occurred more frequently in colon tumours, gliomas, and acute myeloid leukemias (Costello *et al.* 2000).

To summarise, the basis of the argument in this study is as follows: if, as predicted, preferential loss of the active paternal *ZAC* allele is occurring in tumours, those that are shown to have LOH will be expected to show a maternal methylation pattern at NV149, whereas those without LOH would, by definition, retain both parental alleles and would therefore be expected to show both a maternal and a paternal methylation pattern. There is a possibility, however, that *de novo* tumour-specific methylation of NV149 may be occurring. By including the non-LOH tumours in the methylation analysis, this possibility can be assessed – if *de novo* methylation were prevalent, both LOH cases and non-LOH cases would show a maternal methylation pattern at NV149, because this event would be non-discriminating. Thus, the analysis of the non-LOH cases allows distinction between whether a maternal methylation pattern in the LOH cases represents preferential paternal allele loss, or *de novo* methylation in the tumours.

5.2 DEVELOPMENT OF EXPERIMENTAL STRATEGY AND OPTIMISATION OF TECHNIQUES

5.2.1 Selection of microsatellites

Five polymorphic dinucleotide repeat microsatellites were chosen for the purpose of analysing the LOH at *ZAC*/NV149. These were as follows: D6S308, D6S311, D6S279, D6S977, and D6S1703. The markers were selected on the basis of their close proximity to the *ZAC*/NV149 locus, and their high heterozygosity, which maximises the chances of the samples being heterozygous and therefore informative for the markers.

Figure 52 shows the relative locations of the markers. The map positions and genetic distances between markers are according to the LDB database (genetic Location DataBase at the University of Southampton).

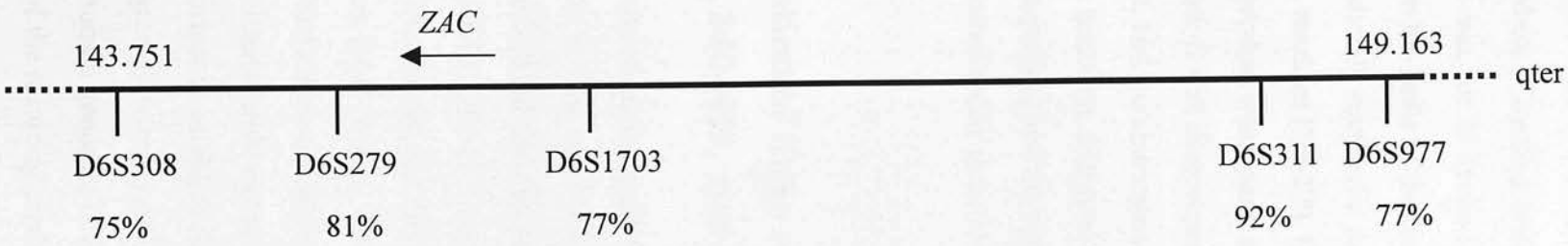


Figure 52. The relative order of the five CA repeat markers chosen for the LOH study, according to the Cedar map. The heterozygosity for each is given as a percentage. Scale is in Mb.

5.2.2 Optimisation of microsatellites

The five microsatellites were tested and optimised first on a panel of normal DNA controls. A standard set of cycling conditions was used initially (e.g. 94°C/1 min; 55°C/1 min; 72°C/2 min, for 35 cycles), and then modified as appropriate for each of the microsatellites on the basis of the results obtained. The markers D6S308 and D6S311 were immediately successful and a satisfactory product was consistently obtained. However, markers D6S279, D6S977 and D6S1703 failed to give an adequate or visible product after many optimisation attempts. On referring back to the original sequences, it was discovered that the published primer sequences were in some cases incorrect, and in other cases perhaps too short to be effective. For these reasons, new primer sets were designed for D6S279, D6S977 and D6S1703 based on available genomic sequence, and the optimisation procedure repeated. This time each set of primers consistently generated a product that was of satisfactory quality for ABI analysis.

5.2.3 Identification of three novel CA repeats in the PAC clones 197-L1, 340-H11, and 468-K18

Whilst the five CA repeat microsatellites described above are all located at 6q23, the closer the markers are to *ZAC* and NV149, the more accurate the results will be as an indication of the true LOH at the *ZAC* locus. For this reason, the sequences of the three PAC clones, 197-L1, 340-H11 and 468-K18, which surround *ZAC*, were searched for novel CA repeats, and this revealed the presence of one such CA/GT repeat in each of these PAC clones. These were named CA197, CA340 and GT468 accordingly, and primers were designed to amplify a region of approximately 250-300 bp around each dinucleotide repeat. Since these three microsatellites were novel markers, it was important to establish firstly whether they were polymorphic, and if so, what was the degree of heterozygosity for each. This was done by amplifying a panel of 27 normal human genomic DNA controls for each microsatellite, and analysing the sizes of the resulting products on ABI gels, as described below. The

results of this analysis are given in Table 9. The heterozygosity values observed are: CA197, 80%; CA340, 58%; GT468, 80%. Clearly, CA197 and GT468 are extremely useful due to their high degree of heterozygosity, because they are likely to be informative in approximately 80% of cases. However, despite the lower heterozygosity of the marker CA340 (58%), this marker may still be useful in conjunction with the others, and is particularly important in this study as it is the closest marker to NV149 and *ZAC*.

Control DNA sample number	CA197 (Hex)	CA340 (Tet)	GT468 (Fam)
1	230/240	247/257	241
2	242/252	245	241/245
3	238/240	245	245/251
4	240/242	245	239/241
5	230/240	245/247	239/241
6	232/244	253/257	237/241
7	230/242	245/247	241
8	234/246	257/259	241/245
9	238/240	247	241
10	230/244	243/255	237/239
11	234/240	245	243/245
12	236	239/245	239/241
13	230/240	247/257	241/245
14	234/240	245	241/245
15	240	247/253	235/239
16	230/240	247/251	241/245
17	-	245	-
18	240	245	239/243
19	240	245	243/245
20	240/242	245	241/245
21	238/240	243/245	243
22	230/244	245/247	237/243
23	234	257	245/249
24	230/248	251/257	241/245
25	240/242	245/247	239/241
26	234/240	247/257	243
Heterozygosity	**(20 of 25) = 80%**	**(15 of 26) = 58%**	**(20 of 25) = 80%**

Table 9. Results of ABI analysis of microsatellite PCR products for the novel markers CA197, CA340, and GT468. DNA sample numbers 1-27 are normal human genomic DNA controls. The sizes of the alleles detected on the ABI are shown in the appropriate columns. Where only one allele size is given, it is assumed that the individual is homozygous for that particular marker. A dash (-) indicates an ambiguous ABI result, or failed PCR. The observed heterozygosities of each marker are given at the end of the table.

5.3.4 ABI 377 analysis of microsatellites

The forward primer of each primer pair was labelled with Hex, Tet or Fam. The PCR products of either two or three different markers could be electrophoresed together in the same lane because they were labelled with different coloured dyes, according to the following groupings:

D6S308 (Tet)
D6S311 (Fam)

D6S279 (Tet)
D6S977 (Hex)
D6S1703 (Fam)

CA197 (Hex)
CA340 (Tet)
GT468 (Fam)

Details of the ABI analysis are given in Materials and Methods. Typically, the PCR products were diluted 1:10 or 1:15, depending on the quality, and 1.5µl loaded.

5.3.5 Optimisation of the bisulphite modification protocol

Bisulphite modification has now been used for a number of years as an effective technique for analysing the methylation status of cytosine residues within CpG dinucleotides (Clark *et al.* 1994). The underlying basis of the technique is the capacity of sodium bisulphite to efficiently convert unmethylated cytosine residues to uracil in single stranded DNA. In the event of a cytosine residue within a CpG dinucleotide being methylated, and existing as 5-methylcytosine, it is resistant to modification, and is not converted to uracil. In effect, therefore, the only remaining cytosines in a DNA strand after bisulphite modification should be those which were methylated in the original sequence.

There are three main steps in the chemical reactions that form the modification process. Figure 53 is a schematic diagram to illustrate the process, which is discussed in more detail by Clark *et al.* (Clark *et al.* 1994). Briefly, the first step involves addition of bisulphite to the 5-6 double bond of cytosine, to form a sulphonated cytosine derivative. The second step is a hydrolytic deamination of cytosine-SO_3, to give uracil-SO_3. The third step is an alkali treatment to remove the sulphonate group, yielding uracil. During the modification of the DNA, unmethylated cytosines will be converted to uracil, whereas methylated cytosines will remain as cytosine. This can be readily detected in the sequence following the modification, by amplifying the region of interest by PCR, during which the uracil residues in the modified sequence are copied to thymine. Thus, the sequence of the PCR product enables the methylation status of C in CpG dinucleotides of the original sequence to be determined, because CpGs that are detected as TpGs in the PCR product will represent unmethylated CpGs in the original sequence, and any remaining CpGs in the PCR product will represent methylated sites.

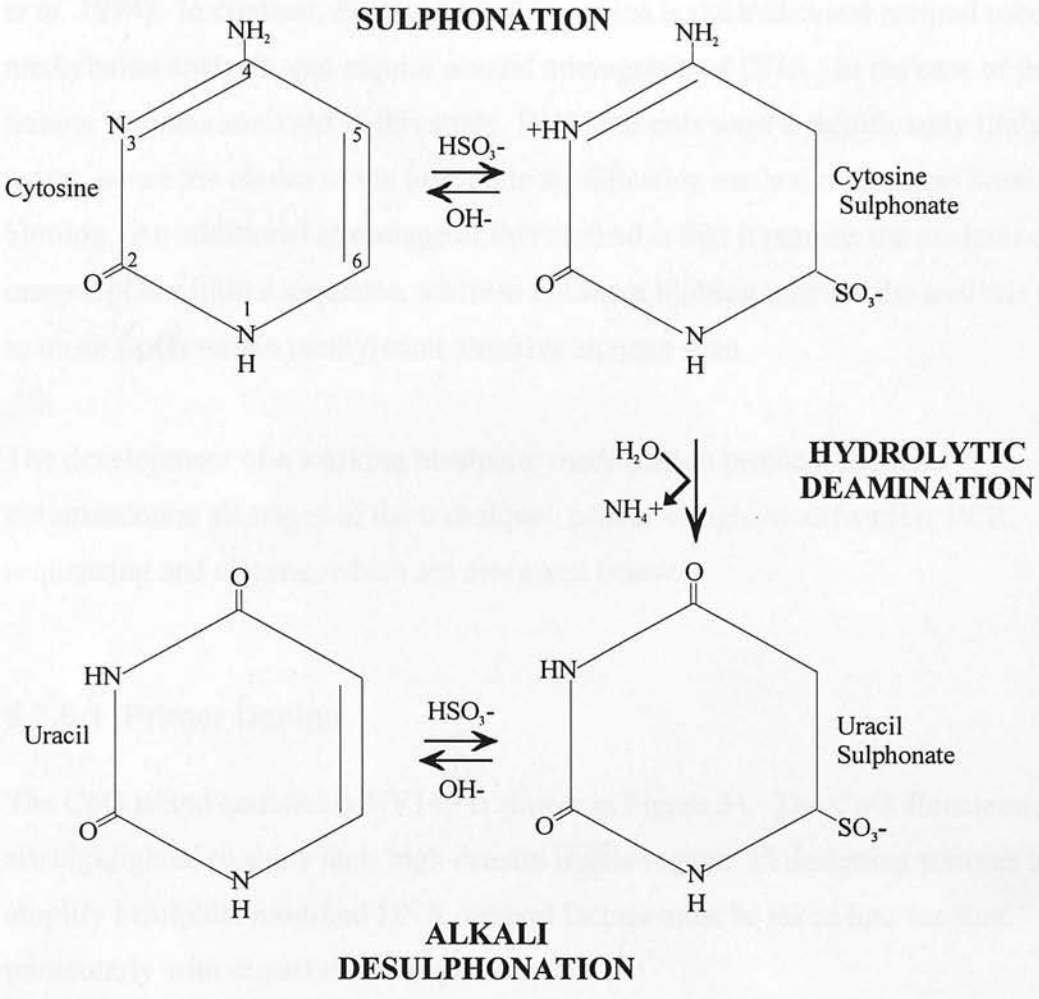


Figure 53. Schematic diagram of the conversion of cytosine to uracil in the bisulphite modification process. Redrawn after Clark *et al.* (Clark *et al.* 1994).

A particularly important advantage of bisulphite modification and sequencing over alternative methods of methylation analysis is that bisulphite modification can be successfully employed with limited amounts of DNA. In practice the technique has been shown to work efficiently on amounts of genomic DNA as low as 200pg (Clark *et al.* 1994). In contrast, Southern blotting, which is the traditional method used for methylation analysis, can require several micrograms of DNA. In the case of the tumour samples analysed in this study, DNA amounts were a significantly limiting factor, hence the choice of the bisulphite modification method rather than Southern blotting. An additional advantage of this method is that it permits the analysis of many CpGs within a sequence, whereas Southern blotting restricts the analysis only to those CpGs within methylation sensitive enzyme sites.

The development of a working bisulphite modification protocol required optimisation at all stages of the technique: primer design, modification, PCR, sequencing and cloning, which are discussed below.

5.3.5.1 Primer Design

The CpG island containing NV149 is shown in Figure 54. The CpG dinucleotides are highlighted to show their high density in this region. In designing primers to amplify bisulphite modified DNA, several factors must be taken into account, particularly with regard to the sequence.

(i) After complete modification, the two strands are no longer complementary to each other. Therefore, unlike a standard PCR, a separate pair of primers would be required to amplify each strand. For the purpose of this experiment, one set of primers was designed with specificity for the top strand only.

(ii) After modification, there are two possible resultant sequences at every position which was a CpG dinucleotide in the original sequence, depending on the methylation status of the cytosine within the CpG, i.e. The cytosine residue within a CpG dinucleotide will remain as cytosine if methylated, but will be converted to

thymine if unmethylated. Unless the methylation status of all CpGs is already known, primers must be designed to accommodate the possibility of a 'C' or a 'T' occurring at every CpG within the primer sequence.

(iii) In general, it is recommended that primers for amplifying bisulphite-modified DNA should be designed slightly longer than typical PCR primers (Clark *et al.* 1994). This is to account for the lower affinity/specificity of the primers for their complementary sequence, which is reduced by the presence of the potentially variable sequence at CpG sites, and by the fact that the conversion of C to T reduces the sequence complexity of the target DNA.

(iv) Smaller regions are more efficiently amplified from modified DNA than larger ones. In this experiment several pairs of primers were designed, to attempt to generate the largest possible product, thus permitting analysis of a greater number of the CpGs within the island. However, a successful PCR product was only obtained using a pair of primers designed by our collaborators (Kamiya *et al.* 2000), which amplified only 154bp of the sequence. The location of these primers (Bssu2 and BSS12) is shown in Figure 54.

(v) The modification process involves an initial digestion with a relatively common restriction enzyme. It is therefore imperative that the region to be amplified does not contain a site for the chosen enzyme between the primers.

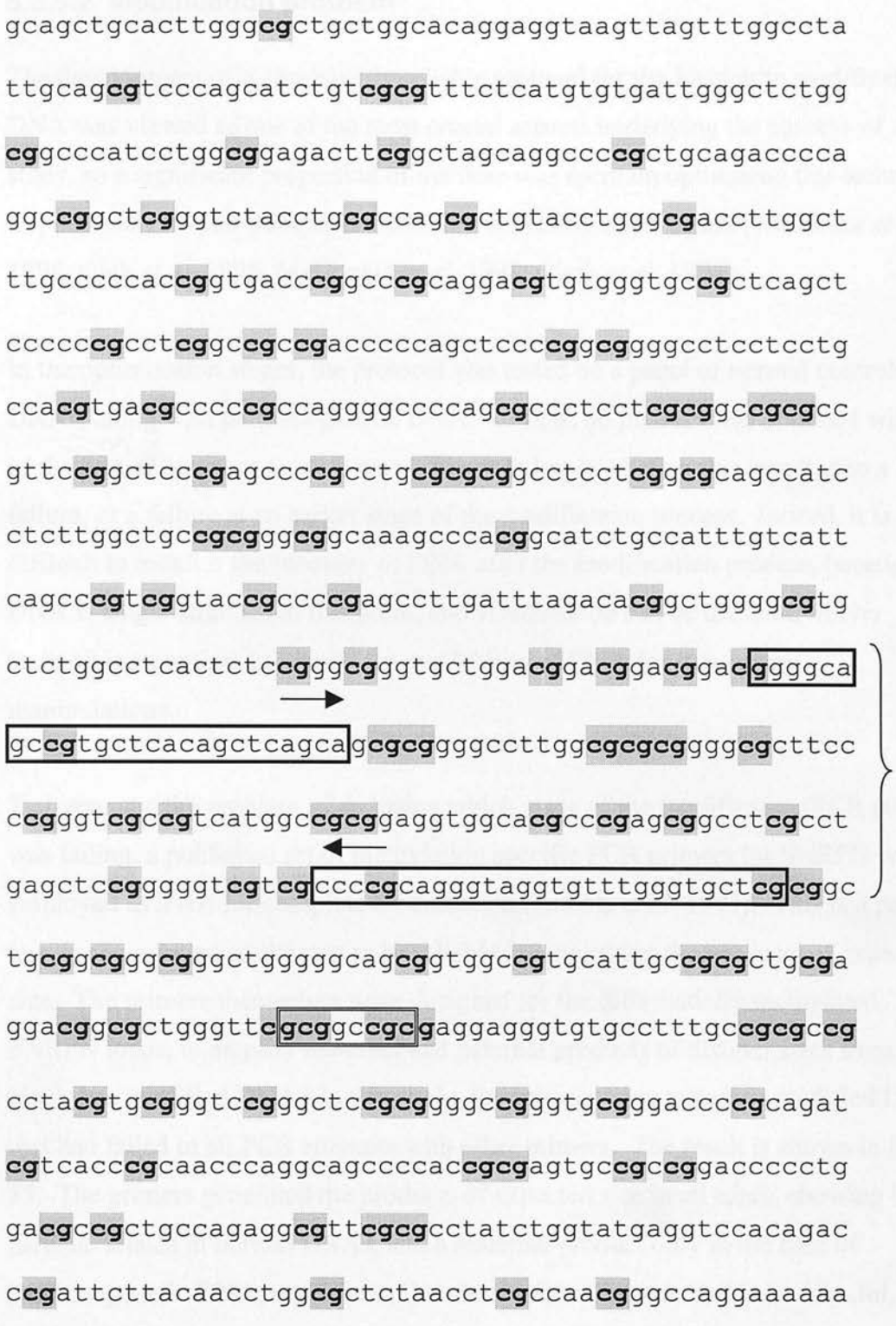


Figure 54. The NV149 CpG island, showing location of the bisulphite PCR primers Bssu2 (forward) and BSS12 (reverse), indicated with boxes, in the directions shown by the arrows. The region amplified by these primers is indicated with a bracket. All CpGs are highlighted, and the *Not*I site of NV149 is shown with a double box.

5.3.5.2 Modification protocol

The development of a consistently reliable protocol for the bisulphite modification of DNA was viewed as one of the most crucial aspects underlying the success of this study, so a significant proportion of the time was spent on optimising this technique. In particular, several published methods were tested and adapted (Warnecke *et al.* 1998; Olek *et al.* 1996; McDonald *et al.* 1998; Clark *et al.* 1994).

In the optimisation stages, the protocol was tested on a panel of normal control DNAs, along with parthenogenetic DNA. At first, no product was obtained with any of the primer sets tested, but it was difficult to know whether this was due to a PCR failure, or a failure at an earlier stage of the modification process. Indeed, it is difficult to monitor the recovery of DNA after the modification process, because the DNA is single-stranded at this point, and visualisation can be difficult. Silver staining is an option, but limits the availability of DNA for subsequent manipulations.

To overcome the problem of deducing which stage of the modification/PCR process was failing, a published set of methylation specific PCR primers for SNRPN were employed as a test for complete modification (Kubota *et al.* 1997). This is a positive test, as the primers are known to be reliable in amplifying the products of expected size. The primers themselves were designed for the differentially methylated SNRPN locus, to amplify maternal and paternal products of distinct sizes from bisulphite modified DNA. In this study, the primers were tested on modified DNA that had failed in all PCR attempts with other primers. The result is shown in Figure 55. The primers generated the products of expected size in all cases, showing both parental alleles in normal DNAs, and a maternal product only in the case of parthenogenetic DNA, confirming that the modification was in fact successful.

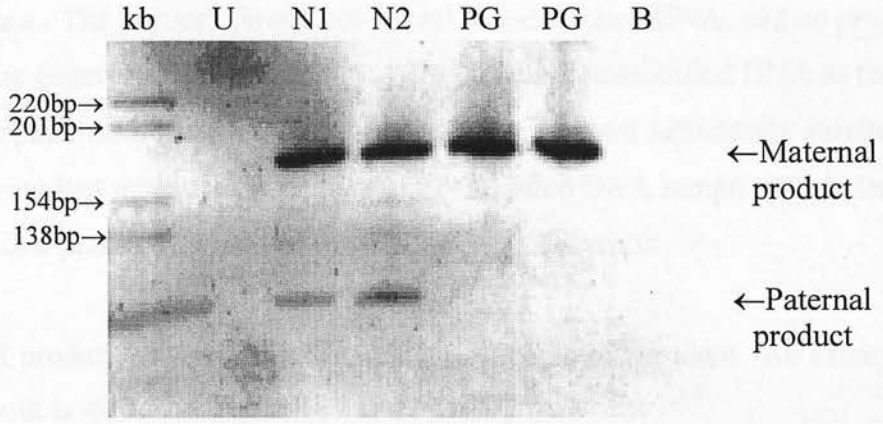


Figure 55. Methylation specific *SNRPN* PCR performed on bisulphite modified normal (N1 and N2) and parthenogenetic (PG) DNAs, as a test for complete modification. A normal sample should generate products of 174bp and 100bp, representing the maternal and paternal alleles respectively. A parthenogenetic sample should generate only the maternal band, as seen here. U is a PCR containing unmodified template DNA; B is a negative control PCR containing no template.

5.3.5.3 PCR on the bisulphite-modified DNA

It was decided to use the primer pair designed for this region in the laboratory of Y. Hayashizaki. The PCR cycling conditions that they had used for this set of primers were tested. The conditions were subsequently modified slightly to optimise the yield and quality of the PCR product. Conditions are given within the modification protocol detailed in Materials and Methods. An important control in this experiment is that a sample of unmodified DNA should be tested with these PCR primers and conditions. The primers should not anneal to unmodified DNA, and no product should be generated. If a product was detected with unmodified DNA as the template, this would suggest that the conditions were not sufficiently stringent, and that the product generated from bisulphite-modified DNA samples might include unmodified products, which would confuse the final result.

No PCR product was amplified from unmodified DNA template. An example of the PCR result is shown in Figure 56.

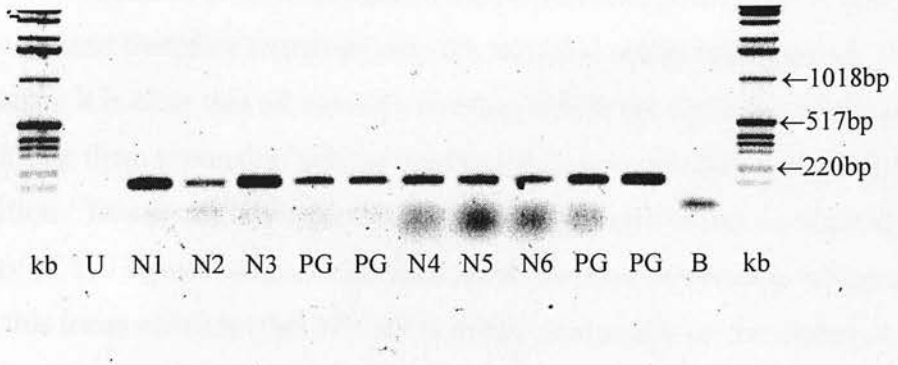


Figure 56. Example of an agarose gel of the bisulphite PCR products generated from some normal control DNAs and parthenogenetic DNA. U is a control PCR containing unmodified normal DNA; N1-N6 are normal DNAs; PG is the parthenogenetic DNA; B is a negative control PCR containing no template. The sizes of bands in the 1kb ladder marker are given on the right.

5.3.5.4 Sequencing the PCR products

The PCR products amplified as described above were sequenced using an internal forward primer. A panel of normal control DNA samples and a parthenogenetic DNA sample were sequenced to establish the normal methylation pattern of the CpGs within the amplified region. The typical result obtained by a standard dGTP sequencing protocol is shown in Figure 57. The sequence on the right is that of a PCR product amplified from bisulphite-modified parthenogenetic DNA, and by definition should therefore represent only the maternal methylation pattern. Within this sequence it is clear that all cytosine residues within the CpGs are methylated, thus affording them protection against conversion to thymine during the bisulphite modification. This result is in agreement with that of experiments conducted in the laboratory of Y. Hayashizaki, and the consistent retention of cytosine within all CpGs at this locus confirms that NV149 is methylated solely on the maternal allele. In addition, this result demonstrates that the modification protocol used here is successful, and complete.

The sequence on the left is that of a PCR product amplified from a normal bisulphite-modified DNA sample, and is somewhat more difficult to interpret. On the basis that the maternal allele is completely methylated, and that there is differential methylation at this locus, a normal DNA sample containing both a paternal and a maternal allele would be expected to show variation at each of the CpG dinucleotides within the bisulphite PCR product. Initially it was anticipated that at each CpG in the sequence, the C would be accompanied by a T of equal intensity (the C representing the methylated maternal allele, and the T representing the unmethylated paternal allele). In practice however, the sequence appears more complex than this, and in fact it appears as if two superimposed sequences are present. It is occasionally possible to detect, as predicted, a C and a T alongside each other in the same position on the gel at a position where the parthenogenetic sequence shows a methylated CpG – one such example is indicated by a box in Figure 57, however, the clarity of the sequence is very poor. More illuminating is the distinct and unexpected pattern present in the A lane of the biparental sample. A comparison of this to the

equivalent lane in the parthenogenetic sample reveals that for every A residue in the parthenogenetic sample, there are two bands in the normal sample. Neither the modified nor the unmodified predicted sequence contains double As, yet there consistently appears to be two A bands at each expected A residue in the normal sample. These are indicated by arrows in Figure 57. The most likely explanation for this is that the sequences of the maternal and paternal alleles, being differentially methylated, deviate considerably from each other in sequence after modification. When run on a gel, the mobilities of the paternal and maternal alleles are sufficiently different from each other to cause the staggered effect seen in Figure 57. The specific cause of the altered mobilities of the sequences is unknown, but may be a result of secondary structure formation, and C/G compressions. These factors certainly play some role, since dITP sequencing was able to resolve the effect to some extent.

Sequencing directly on the bisulphite PCR products is useful as a preliminary indication of the methylation status of the sample, but in the case of a differentially methylated sequence such as this, the resultant sequence comprises a mixture of paternally- and maternally-derived products. The intended strategy in this experiment was to assess the methylation status (and thus determine the parental origin) of the remaining allele in tumours where loss of one allele had previously occurred by deletion. Clearly, the successful excision of most tumours results in the unavoidable presence of some surrounding normal cells in the sample. This contamination of the tumour sample with normal material can significantly affect the results, unless the experimental design accounts for the effect. A potential solution to the problem of such contamination, and to the sequencing problem described above is to clone the PCR products and sequence a selection of the resulting clones, as discussed below.

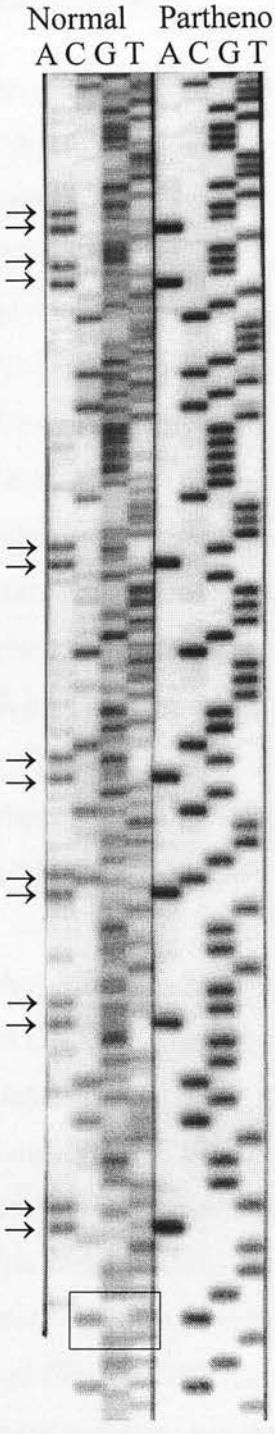


Figure 57. The result of sequencing the bisulphite PCR products amplified from normal and parthenogenetic (Partheno.) bisulphite-modified DNAs. The arrows and the box indicate distinctive features of the sequences, which are discussed in the text, and are used subsequently in classifying the methylation patterns of tumour samples.

5.3.5.5 Cloning the PCR products

For the purpose of this experiment, the plan was to clone each of the PCR products amplified from bisulphite-modified tumour DNA (and normal DNA where available), and sequence approximately 20-30 clones in each case. The nature of the cloning system means that each clone will be either maternal or paternal in origin, this being determined from the sequence at each CpG as described above. The cloning of a bisulphite PCR product derived from normal DNA should be expected to generate on average 50% paternal clones and 50% maternal clones, and a result close to this ratio should be observed if sufficient clones are analysed. In analysing the tumour samples, the hypothesis is that a tumour with LOH retains only one allele, and if *ZAC* plays a suppressor role, showing that the retained allele is always the inactive, methylated, maternal copy will substantiate this idea. By this hypothesis, cloning the bisulphite PCR product of a tumour sample with LOH should therefore generate clones which are all maternal. However, as explained, the contamination of the sample with a small proportion of normal tissue may affect this, so that a small number of paternal clones may be identified.

To develop and optimise the cloning part of the protocol, the PCR products of normal bisulphite modified DNA were tested. The PCR product was used in the ligation reaction directly; ligation into the pGEM®-T Easy vector, and transformation into JM109 cells was performed essentially as described in the Promega protocol, and the cells were plated out onto ampicillin selective LB plates, also containing X-gal and IPTG (this procedure is detailed in full in Materials and Methods). On average, around 200 colonies were observed on the plates after overnight incubation, and of these, 20 were picked for miniprep. Colony PCR confirmed that they all contained an insert of the expected size, and thus there was no evidence of the cloning of primer-dimer bands. Miniprep extractions were performed using the SNAP™ miniprep kit, and the inserts sequenced using the M13 forward primer within the vector sequence.

Unexpectedly, the sequencing of 20 clones that originated from the normal control DNA, revealed that only two were maternal in origin, with the remainder being paternal. Though statistically unlikely, this result is possible with a relatively small sample. However, repeating the PCR and the cloning procedures as before generated a similar ratio of three maternal clones to 17 paternal clones, suggesting that there is a consistent cloning bias. The reason for this is unknown, but the fact that the cloning system has a preference for paternal over maternal PCR products renders it useless as a means of determining the ratio of maternal and paternal alleles present in the original sample.

5.3.5.6 Final strategy for analysis of breast tumours, bladder tumours and lymphomas

As a result of the procedures described above, the strategy for the analysis was finalised as follows:

Microsatellite analysis to be performed as described, to classify the tumours into two subgroups, comprising those showing LOH at the markers tested, and those not showing any LOH.

Bisulphite modification, PCR and direct sequencing (dGTP) of PCR products to be carried out on all normal and tumour samples. Cloning of the PCR products is eliminated from the final strategy due to the observed bias for paternal products.

Bisulphite sequencing gels to be analysed essentially by a comparison with the normal and parthenogenetic controls, with particular reference to the A residues, since the two parental alleles have been found to migrate differently on the gel in a consistent and reproducible manner. Therefore, the presence of a double A band on the gel at each adenine position indicates a biparental methylation pattern. The presence of only the lower A band indicates a maternal methylation pattern only. Consistent differences in intensity between the upper (paternal) and lower (maternal)

bands of each doublet might also serve to indicate incomplete loss of either allele or changes in methylation at this locus.

5.3 RESULTS – BREAST TUMOURS AND AND BLADDER TUMOURS

5.3.1 The panel of tumour DNAs

A panel of 11 paired breast tumour/normal DNAs and 8 paired bladder tumour/normal DNAs was obtained from Dr. A. McCann, University College Dublin. Approximately 2μg of each was available, of which 1μg was required for the bisulphite modification analysis. The remaining 1μg was used for the microsatellite PCRs.

5.3.2 Microsatellite analysis

The tumours and their corresponding normal DNAs were analysed with the five microsatellites D6S279, D6S308, D6S311, D6S977, and D6S1703. The three novel markers CA197, CA340 and GT468 were not available at this time, and there was insufficient DNA remaining from these tumours when they later became available, so only the original five markers were included in this analysis.

The results of the analysis are summarised in Table 10. Loss of heterozygosity was apparent in several cases, as judged by the reduced peak height of one of the alleles in the tumour DNA relative to the normal DNA control. However, the results were found to be distorted by the apparent contamination of tumour DNA with normal DNA. This proved to be a significant problem, since LOH was only barely detectable in some cases.

One breast tumour showed some loss of heterozygosity at three of the five markers, with the remaining two markers being uninformative. These results for this tumour/normal DNA pair (N94/T94) are shown in Figure 58. The 'diagnosis' of LOH in this case is based on a visual assessment of the peak heights. A numerical analysis may be more appropriate, although the peak heights vary greatly between

markers and between individual PCR products depending on the quality of each PCR product.

In summary, judging by the reduced peak heights, five of eight bladder tumours (62.5%) showed some degree of loss for one allele of at least one of the markers tested. Three of eleven breast tumours (27%) showed similar loss for at least one marker. However, the problem of contamination of these tumour samples with surrounding normal tissue means that the capacity for detection of LOH is less than satisfactory.

Tumour/normal DNA pair	D6S308	D6S279	D6S1703	D6S311	D6S977
N1	199	287	218	226/236	228/232
T1	199	287	218	226/236	228/232
N7	193/199	291/295	214/218	226/236	226/232
T7	193/199	291/295	214/218	226/236	226/232
N12	201	285	218/222	236/256	224/232
T12	201	285	218/222	236/256	224/232
N32	193/203	285	216/222	224/238	222/232
T32	193/203	285	216/222	224/238	222/232
N53	199/201	285	216/218	220/238	222/226
T53	199/201	285	216/218	220/238	222/226
N65	197/203	287/293	218	240	230/232
T65	197/203	287/293	218	240	230/232
N73	193/199	291	216/222	254/262	222
T73	193/199	291	216/222	254/262	222
N82	199	291	216/220	234/236	222
T82	199	291	216/220	234/236	222
N84	193/201	289	214/218	223/252	222
T84	193/201	289	214/218	223/252	222
N94	193/199	287	216	254/256	226/230
T94	193/199	287	216	254/256	226/230
N95	199	291/293	218/224	236/258	222/232
T95	199	291/293	218/224	236/258	222/232

Table 10. Results of ABI analysis of microsatellite PCR products for paired breast normal and tumour DNAs. In cases where there was apparent partial loss of one allele, the affected allele is highlighted.

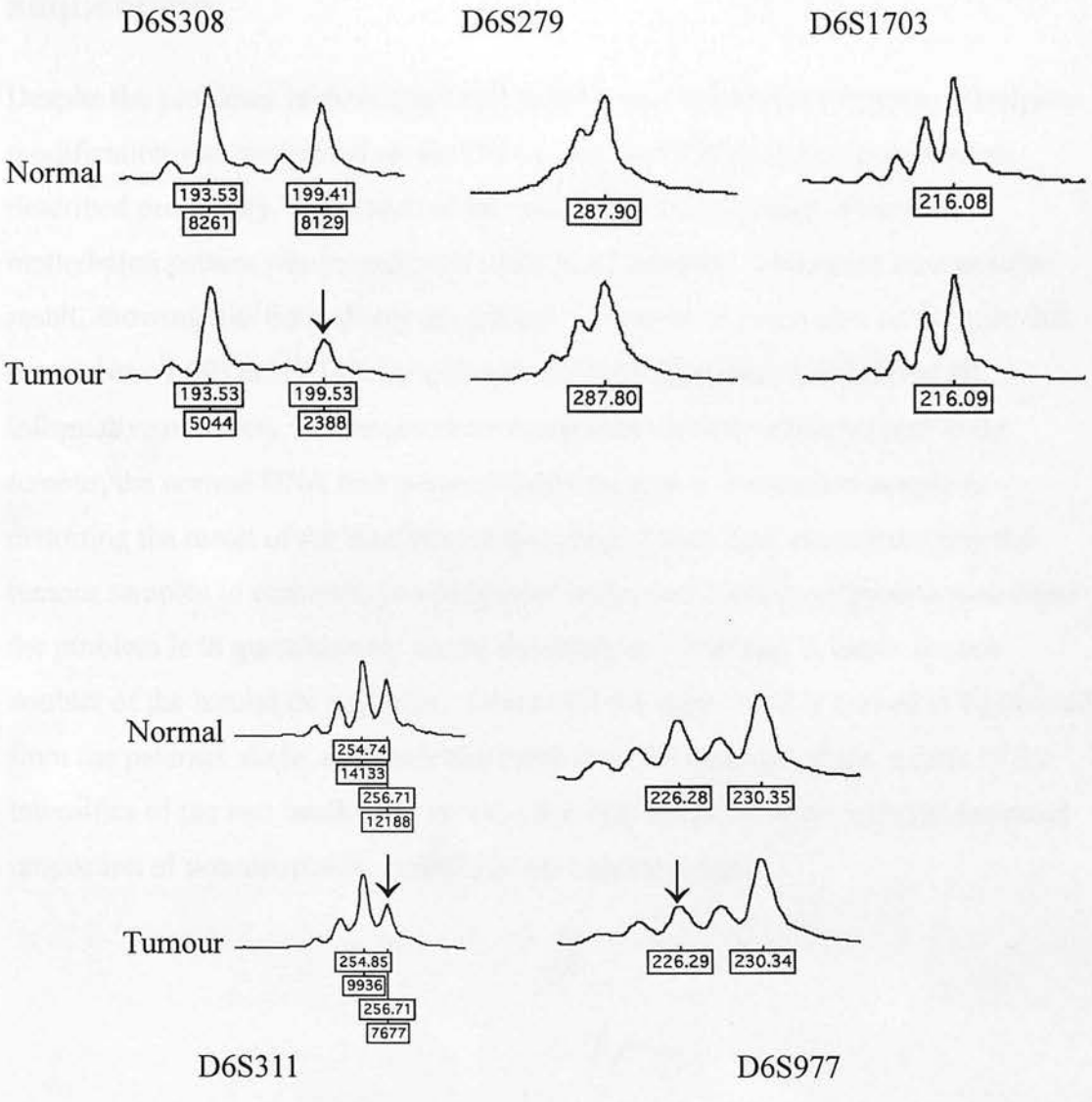


Figure 58. Example of the LOH analysis for normal/breast tumour pair 94. In this case, there is apparent LOH for three of the markers (D6S308, D6S311 and D6S977) in the tumour; the lost allele is shown with an arrow. The markers D6S279 and D6S1703 are non-informative. The boxes under the peaks show the allele sizes; where there are two boxes, the lower box in each case shows the peak area.

5.3.3 Analysis of methylation by bisulphite modification and sequencing

Despite the problems of detecting LOH in the breast and bladder tumours, bisulphite modification was performed on the DNAs, and the PCR products sequenced as described previously. The result of this suggested that a normal, biparental methylation pattern was present at NV149 in all samples. This could be a genuine result, showing that both alleles are present. However, it could also be the case that there is true LOH at NV149, for example in N94/T94 (which had LOH at all informative markers), but despite there being only one true allele present in the tumour, the normal DNA that is unavoidably present in the tumour sample is distorting the result of the bisulphite sequencing. Other than microdissecting the tumour samples to eliminate non-neoplastic cells, one possible solution to overcome the problem is to quantitatively assess the intensity of the two A bands in each doublet of the bisulphite sequence. Given that the upper band is known to be derived from the paternal allele, and the lower band from the maternal allele, a ratio of the intensities of the two bands may reveal a bias that could correlate with the expected proportion of non-neoplastic material in the tumour sample.

5.4 RESULTS – BREAST CANCER CELL LINES

5.4.1 The breast cancer cell lines

Six breast cancer cell lines were obtained from Dr. V. Speirs, Molecular Medicine Unit, St. James's Hospital, Leeds. These were MDA-MB-435, MDA-MB-468, T47D, MCF7, MCF7-Clone 9, and ADR-MCF7. Unlike the breast and bladder tumours analysed previously, this panel of breast cancer cell lines (and the panel of lymphomas – see 5.5) has no matched normal DNA for use as a control. Despite this, it was decided that they could still be analysed in a similar way to the breast and bladder tumours, with one main difference in the LOH experiments. Clearly, without a normal DNA control, it is impossible to say for certain that loss of heterozygosity has occurred solely in the tumour/cell line, yet if apparent homozygosity was observed at several of the microsatellite markers, one could make an assumption about the possibility of this representing LOH, based on the statistical unlikeliness of this multiple homozygosity occurring by chance. This type of analysis has been described and utilised by Goldberg *et al.* (Goldberg *et al.* 2000), in order to identify regions of hemizygous deletion in melanoma cell lines lacking matched controls. They designated the method as HOMOD, (homozygosity mapping of deletions), and established a significance level of 0.001, below which they would assume hemizygosity rather than homozygosity. To attain this level of statistical significance, apparent homozygosity (implicated by a single allele) at five or more adjacent markers was usually required. By this method, Goldberg *et al.* detected deletions in 20 of 40 cell lines, represented by homozygosity of between 7 and 30 adjacent markers. In conjunction with standard LOH results of matched normal and tumour pairs, the results of the HOMOD analysis of cell lines enabled them to define six small consensus regions of deletion, corresponding to putative melanoma tumour suppressor genes on chromosome 11.

In the context of this study, if a breast cancer cell line (or lymphoma sample, see 5.5) was found to contain only one allele at every one of the eight microsatellite markers

analysed, the probability of that cell line being genuinely homozygous with two identically-sized alleles for all markers can be calculated as follows:

Referring back to the heterozygosity values for each marker, the homozygosity expected for each is:

D6S308	25%
D6S279	19%
D6S1703	23%
D6S311	8%
D6S977	23%
CA197	20%
CA340	42%
GT468	20%

Probability of being genuinely homozygous at all eight markers, assuming no linkage disequilibrium= $0.25 \times 0.19 \times 0.23 \times 0.08 \times 0.23 \times 0.20 \times 0.42 \times 0.20$

$= 3.38 \times 10^{-6}$, 0.000338%, or 1 in 263157.

Clearly therefore, a finding of apparent homozygosity at all eight markers would be strongly suggestive of hemizygosity or deletion in the cell line/tumour.

Similarly, homozygosity for only the first five of the markers would still be statistically unlikely at 1 in 4974, and probabilities of other combinations of apparent homozygosity can be calculated accordingly.

Analysis of the both the breast cancer cell lines and the panel of lymphomas was continued with on the basis of this assumption, so that those showing apparent homozygosity at multiple adjacent markers could be 'diagnosed' as cases with deletion at this chromosomal region.

5.4.2 Microsatellite analysis

The analysis of LOH in the cell lines was performed essentially as with the breast and bladder tumours, but as explained above, LOH was judged by apparent homozygosity at multiple adjacent markers. The results are shown in Table 11. Of the six breast cancer cell lines, one (ADR-MCF7) was found to have only a single allele for all eight markers analysed. The probability of this representing true homozygosity rather than hemizygosity is 1 in 263157, suggesting that it is more probable that this cell line has a deletion across this chromosomal region. Other than this, the cell line MDA-MB-468 is apparently homozygous for the three adjacent markers D6S1703, D6S311 and D6S977, which may represent hemizygosity for this region, but is not significant in the context of loss of *ZAC*, because this cell line is clearly heterozygous for the three markers that are closest to *ZAC*/NV149 (CA197, CA340, GT468). In summary therefore, it is likely that only the cell line ADR-MCF is hemizygous for the chromosomal interval from D6S308 to D6S977 (containing *ZAC* and NV149). This hemizygosity could be due to a deletion, or loss of the whole chromosome.

Cell Line	D6S308	D6S279	GT468	CA340	CA197	D6S1703	D6S311	D6S977
MDA-MB-435	193	285/287	243/245	245/247	238/240	218/220	262/262	222/224
MDA-MB-468	197	287/301	227/241	251/255	230/242	218	229	222
T47D	197/203	291	247/251	243/245	-	208/218	227/233	220
ADR-MCF7	193	270	245	255	240	226	260	232
MCF7	193/226	285/295	237/243	247/255	230/242	216/220	226/236	222/230
MCF7-Clone9	193/226	285/295	237/243	247/255	230/242	216/220	226/236	222/230

Table 11. Results of ABI analysis of microsatellite PCR products for six breast cancer cell lines. The marker order shown here is the same as their chromosomal order, centromere to telomere. CA340 is the nearest marker to *ZAC*. Where only one allele was detected, the box is shaded, to illustrate possible regions of hemizygosity. A dash (-) indicates an ambiguous ABI result or failed PCR.

5.4.3 Analysis of methylation by bisulphite modification and sequencing

By the hypothesis stated earlier in this chapter, if *ZAC* is a tumour suppressor gene involved in the malignancies investigated here, a cell line (or tumour) demonstrating LOH at the *ZAC* locus might be predicted to have lost the active paternal allele of the gene, thus retaining only the silent, non-functional copy. The bisulphite analysis of the differentially methylated NV149 CpG island enables distinction between the parental alleles present in the sample on the basis of their methylation status, given that the maternal allele is known to be methylated and the paternal allele unmethylated. Bisulphite analysis was performed on the cell lines, and the PCR products generated from the modified DNA samples were sequenced as described earlier. The results are shown in Figure 59. Given that the cell line ADR-MCF7 was shown to carry a deletion of the *ZAC* locus, it might be expected that the remaining allele in this case would be the inactive maternal (methylated) copy. However, the result of the bisulphite analysis suggests that this is not the case, and in fact the methylation pattern of this cell line indicates that there is a predominance of the paternal (unmethylated) pattern. Additionally, whilst it is very clear from the LOH analysis that ADR-MCF7 carries only one allele for this region, and seemingly it is the maternal rather than paternal allele that is deleted, there is some evidence from the sequencing gel of a methylated allele. If the cell line is deleted, and there is only one allele present, then the presence of a methylation pattern reminiscent of both parental alleles suggests that some subsequent methylation of a subset of the paternal alleles has occurred.

The analysis of MDA-MB-468, MCF7 and MCF7-Clone 9 appears normal in that both methylated (maternal) and unmethylated (paternal) alleles are present, as predicted by the normal biparental results of the LOH analysis which showed no evidence for deletion. The result of the cell line MDA-MB-435 is not clear. Of particular interest however is the result of the cell line T47D. This cell line also appeared to have a normal biparental contribution, as judged by the LOH analysis, yet the bisulphite sequencing shows that there is a predominantly maternal

methylated pattern, similar to the result obtained from the parthenogenetic sample. The paternal allele (upper A band in each doublet) is still faintly visible, but it would seem that there is a predominance of maternal-typical alleles (lower A band in each doublet). This is a significant result because there is no evidence for deletion in this case. Whilst it cannot be ruled out that there is an undetected microdeletion, it is possible that this result is caused by hypermethylation of the NV149 locus, such that the majority of alleles adopt a maternal-type methylated status. This hypermethylation could be predicted to silence the *ZAC* gene, and may therefore represent an alternative means, other than deletion, by which the function of the gene is ablated.

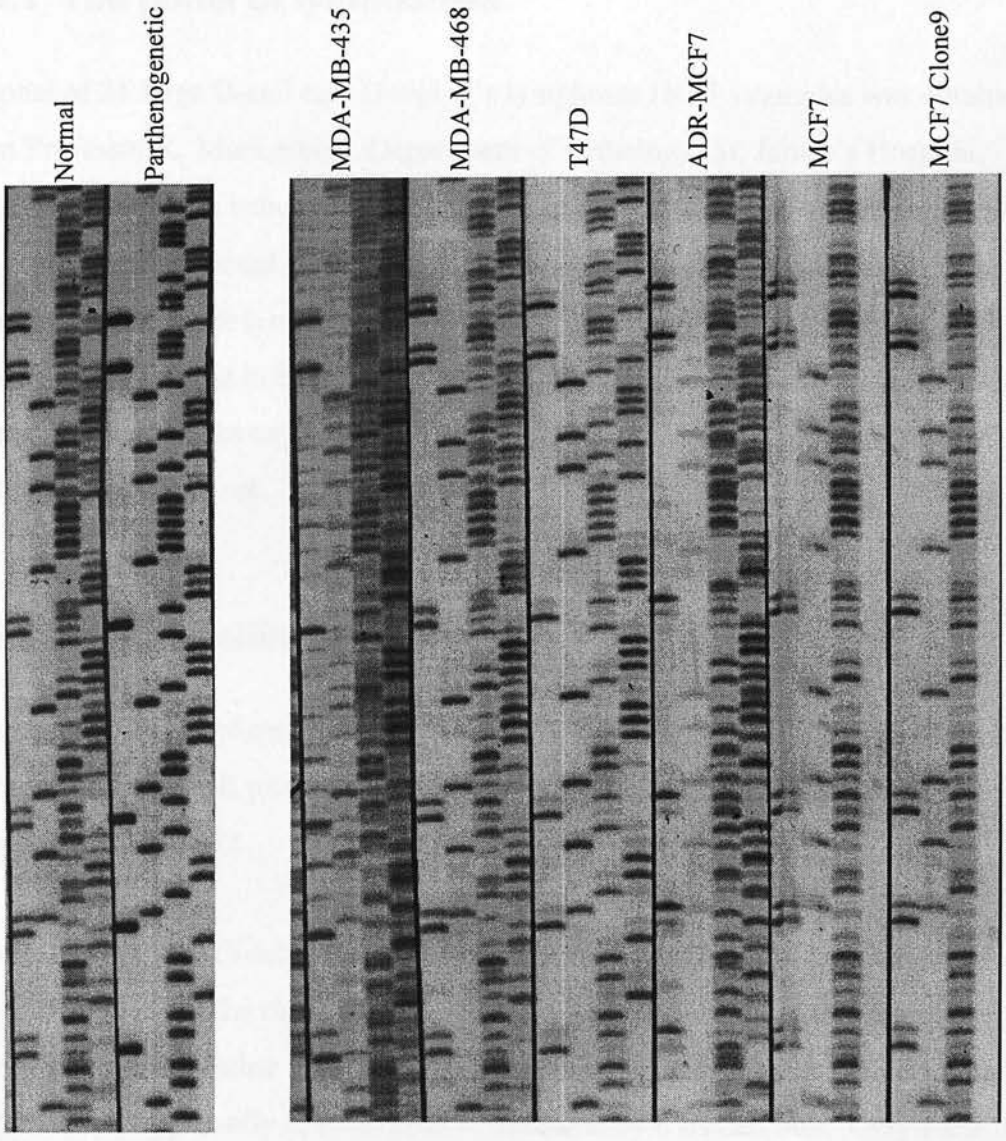


Figure 59. Bisulphite sequencing for the breast cancer cell lines. A normal sample and the parthenogenetic sample are shown on the left for the purpose of comparison. The order of the lanes in each sequence is A, C, G, T. The cell line T47D has an apparent predominantly maternal methylation pattern, whereas ADR-MCF7 has a predominantly paternal methylation pattern, as judged by the doublet A bands in which the upper band is derived from an unmethylated allele, and the lower band is derived from a methylated allele. See text for discussion.

5.5 RESULTS – LYMPHOMAS

5.5.1 The panel of lymphomas

A panel of 21 large B-cell non-Hodgkin's lymphoma (NHL) samples was obtained from Professor K. MacLennan, Department of Pathology, St. James's Hospital, Leeds. The samples were excised tissues, frozen at -70°C for up to five years. DNA and RNA were extracted from these as described in Materials and Methods. The tumour content of the lymphomas in this panel is generally known from histochemical criteria to be approximately 80%, indicating that contamination with normal DNA could be expected to pose less of a problem for LOH analyses than with the breast tumours.

5.5.2 Microsatellite analysis

The panel of 21 lymphoma DNAs was amplified for each of the eight microsatellite markers, and the PCR products analysed on an ABI 377. The results of the analysis are shown in Table 12.

As discussed above (5.4.1), LOH could be assumed if a lymphoma showed only one allele for several of the eight markers, such that the probability of this representing true homozygosity rather than hemizygosity would be very low. In Table 12, each marker that is apparently homozygous is indicated by a shaded box. Only three of the lymphomas displayed homozygosity for four or more markers: L/05, L/14, and L/21. However, these are unlikely to represent true deletions, since the markers that are homozygous are not contiguous. In any case, the three markers nearest to *ZAC* and NV149 (CA197, CA340 and GT468) are the most significant loci at which to find deletion, and only one lymphoma (L/21) is homozygous for two of these. All other lymphomas were heterozygous for two or three of these three markers. Overall, it is therefore unlikely that any of the lymphomas have genuine LOH at the NV149/*ZAC* locus. To find such a lack of LOH in this panel of lymphomas is

somewhat surprising, because it is evident from the published literature that an average LOH of 30-40% could be expected in non-Hodgkin's lymphoma. This result could be due to chance, but it is perhaps more likely that some instances of LOH have been missed because of the lack of matched normal samples for comparison. Indeed, where a lymphoma sample was apparently heterozygous for a marker, a true hemizygous result may have been obscured by the presence of the two original alleles from non-neoplastic cells.

Lymphoma Code	D6S308	D6S279	GT468	CA340	CA197	D6S1703	D6S311	D6S977
L/01	193/199	273/285	243/245	245/261	234/238	218/222	238/250	222/226
L/02	193/201	271/285	241/245	243/245	238/243	218/224	236/238	222/226
L/03	193/199	271/299	245/249	251/255	238/240	216/222	250/252	222
L/04	201	271/285	243/247	245	234/238	216/220	258/264	224
L/05	199	287/291	241/245	251/257	234/240	218	258	228
L/06	197/199	287	243/245	243/249	240	216/218	226/238	224
L/07	193	271/285	241/243	253/255	242/246	216/218	220/226	224/228
L/08	199	285/291	243/245	243/253	238/254	218/220	236/242	222
L/09	199	271/293	241/245	241/245	230/238	218/220	224/240	226/232
L/10	199	287/299	245/249	247/253	230/238	216/220	236	224/228
L/11	193/199	271/295	239/245	253/257	234/242	218	254/260	222
L/12	199	271/289	239/241	241/249	238/240	216/220	224/228	224
L/13	193/195	271/285	247	243/253	240/242	220	236/240	224/232
L/14	199	271	237/243	243	230/236	218	256/260	230/232
L/15	199	287/303	241/245	243/253	242	216/220	226/240	222/226
L/16	199/201	271/287	239/241	243/255	240/242	214	236/256	222/232
L/17	193/199	271	243/249	247/255	234/238	-	238/254	222/226
L/18	193/195	285/297	241/249	253	238/244	216/224	232/238	222/228
L/19	193/203	291/301	245	239/257	234/238	218	236/240	226/228
L/20	195	285/295	247/249	243/255	240	218/222	220/250	222/224
L/21	199	-	241	243	244/258	218	226	222/230

Table 12. Results of ABI analysis of microsatellite PCR products for lymphoma DNA samples 1-21. The marker order shown here is the same as their chromosomal order, centromere to telomere. CA340 is the nearest marker to *ZAC*. Where only one allele was detected, the box is shaded, to illustrate possible regions of hemizygosity. A dash (-) indicates an ambiguous result or failed PCR.

5.5.3 Analysis of methylation by bisulphite modification and sequencing

Despite the apparent lack of LOH in the lymphoma DNAs, all samples (L/01-L/21) were subjected to bisulphite modification and PCR, and then the PCR products sequenced as described previously. The lack of normal DNA for comparison demanded therefore that the lymphomas be closely scrutinised against a bisulphite-modified normal and parthenogenetic DNA control, in order to establish the parental origin of the methylation pattern at NV149.

The lack of LOH in the lymphomas indicates that both alleles are present at the NV149 CpG island. Since both alleles are present, the initial hypothesis would predict a biparental methylation pattern at NV149, displaying both the methylated maternal allele and the unmethylated paternal allele. Results of the bisulphite sequencing for lymphomas L/01-L/08 are shown in Figure 60. Whilst the majority of the sequences show the typical pattern, represented by two superimposed sequences, with a double A band at each A site in the original sequence, there are two lymphomas which clearly deviate from this expected result. Lymphoma sample L/06 appears completely methylated, as the sequence appears identical to that of the parthenogenetic DNA control. Only the lower A band representing the maternal allele is present, suggesting that both alleles are methylated. There is no evidence of any unmethylated sequence in this sample. The absence of LOH, in conjunction with the complete methylation observed at NV149 in this lymphoma, suggests that NV149 has become hypermethylated. Because methylation is normally associated with the silent allele of *ZAC*, the occurrence of *de novo* methylation in the tumour may be hypothesised to result in ablation of function, since both alleles are now effectively silenced.

The lymphoma sample L/05 also appears to have undergone hypermethylation at NV149, since the bisulphite sequence for this sample too shows a predominantly maternal methylation pattern. This sample does have a low, but detectable level of unmethylated sequence. This is quite likely, however to represent the small

proportion of non-neoplastic cells within the lymphoma samples. Histological examination shows that, in contrast to the breast and bladder tumours, these samples of large B-cell lymphoma are comparatively homogeneous, consisting of at least 80% tumour cells.

Thus, although this was not our original expectation, it appears that as with the breast cancer cell line T47D, lymphomas L/05 and L/06 have hypermethylation at the NV149 CpG island, and that this event could play a significant role in silencing both copies of *ZAC* in these patients, thus causing a loss of function. To pursue this idea, it would be necessary to show that the hypermethylation of the NV149 CpG island in these cases is associated with a lack of expression at the mRNA level. This would strengthen the postulate that hypermethylation causes a reduction or complete loss of *ZAC* mRNA, and that the absence of the resulting protein could have a pathological outcome. The technique of RNase protection (RPA) was chosen to assess expression of *ZAC* in the lymphomas, as described below.

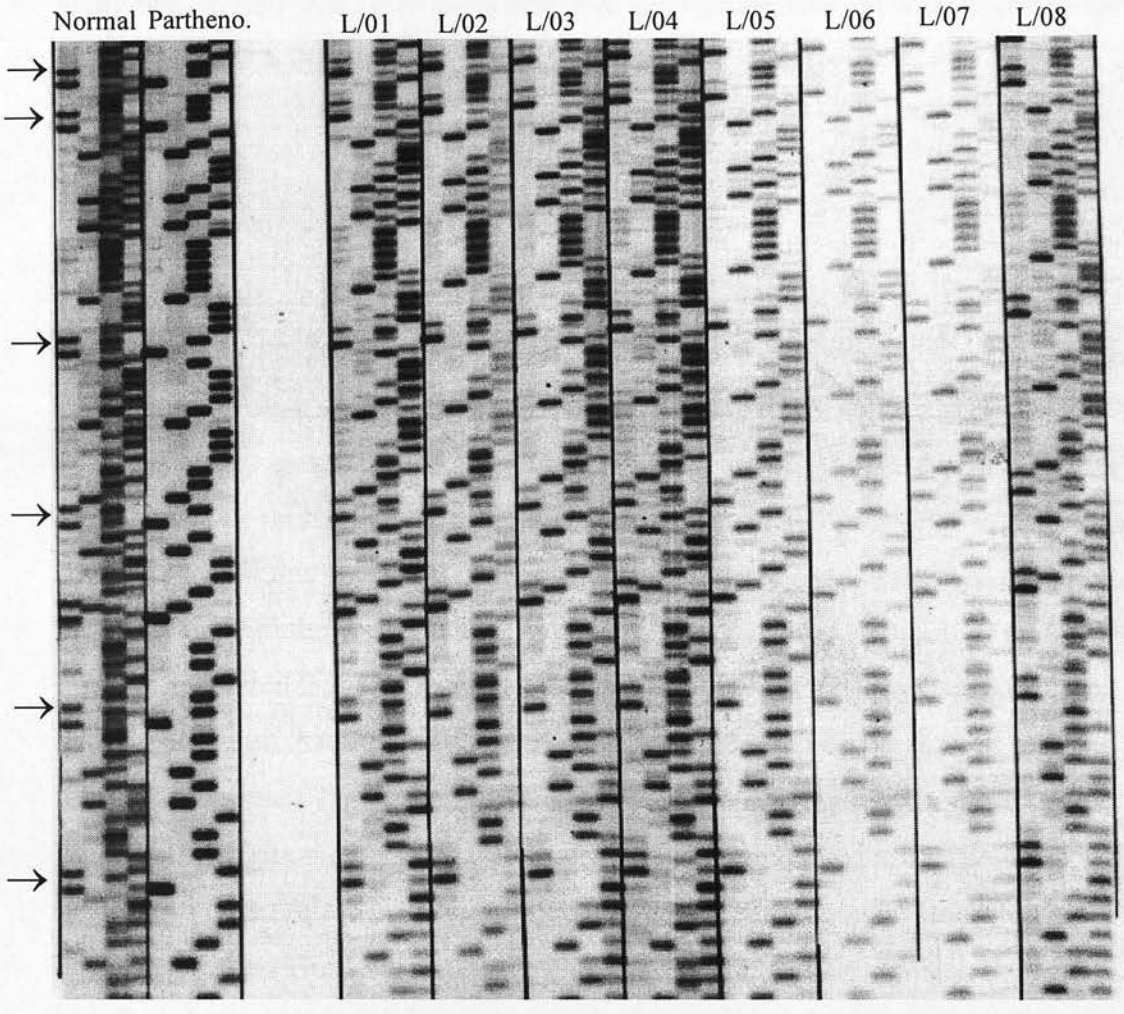


Figure 60. Analysis of NV149 methylation status by bisulphite sequencing. The normal DNA and parthenogenetic DNA references are shown on the left of the gel. Samples L/01 to L/08 are the bisulphite PCR products of lymphoma DNAs L/01 to L/08. Lymphoma L/06 shows a completely methylated pattern, identical to that of the parthenogenetic reference. Lymphoma L/05 also shows a similar pattern, but with a limited degree of paternal contribution. The arrows indicate the A doublets, as discussed in the text.

5.5.4 Ribonuclease Protection Assay (RPA)

For the quantification of *ZAC* levels in the lymphomas, a ribonuclease protection assay was chosen, since it is more sensitive and less demanding with regard to the quality of sample RNA than the Northern blotting technique.

5.5.4.1 Background information

The ribonuclease protection assay (RPA) is a sensitive method for detecting and quantifying a specific target RNA fragment in a sample of total RNA. The technique involves generating a radioactively labelled antisense RNA probe that is complementary to the target RNA. The probe is synthesised such that it is longer than the target; this is usually achieved by cloning the target sequence into a vector (pGEM Easy in this case), and then transcribing *in vitro* from one of the polymerase promoters within the vector (either directly on the linearised plasmid, or on a PCR product generated from it), such that the probe comprises the target sequence, and an additional region that represents the vector sequence between the cloning site and the relevant promoter. The probe is then hybridised to the sample RNA, followed by digestion with ribonuclease, which degrades any unhybridised probe, and also degrades the part of the probe described above that is not complementary to the target sequence. Thus, when electrophoresed on a polyacrylamide gel, the full-length probe is distinguishable in size from the protected fragment that will remain after the digestion. The protected fragment can be visualised by autoradiography, and represents the hybrid of probe and target. Provided there is a molar excess of probe relative to target in the hybridisation reaction, the intensity of the protected fragment is directly proportional to the amount of target RNA in the sample, and thus allows accurate quantification.

5.5.4.2 Generating a *ZAC* antisense RNA probe

Figure 61 is a schematic diagram to explain how an antisense RNA probe was generated for *ZAC*. The RT-PCR product 1029 (see 4.2.17) of a fetal heart cDNA

was cloned into the vector pGEM®-T Easy, and DNA was prepared from six of the resulting colonies. PCR was performed with primers M13F and ZACPROBE5' to identify a clone in the correct orientation such that transcription with T7 RNA polymerase would generate an antisense rather than sense probe. These primers should generate a product only from those clones that are in the required orientation; Figure 62 shows the result, which indicates that clones 2, 5 and 6 conformed to this requirement. Clone 2 was selected for further procedures. The PCR product described above was prepared for use as a template in the transcription reaction, by treating with SDS and proteinase K to remove any nucleases, and then purifying with a Wizard prep column. For transcription, 100ng was used. The transcription reaction was performed as recommended in the Ambion protocol (see Materials and Methods). The amount of limiting nucleotide was optimised first in an unlabelled reaction, and a final concentration of 4μM limiting nucleotide was chosen as the optimum, in order to generate a probe of high specific activity.

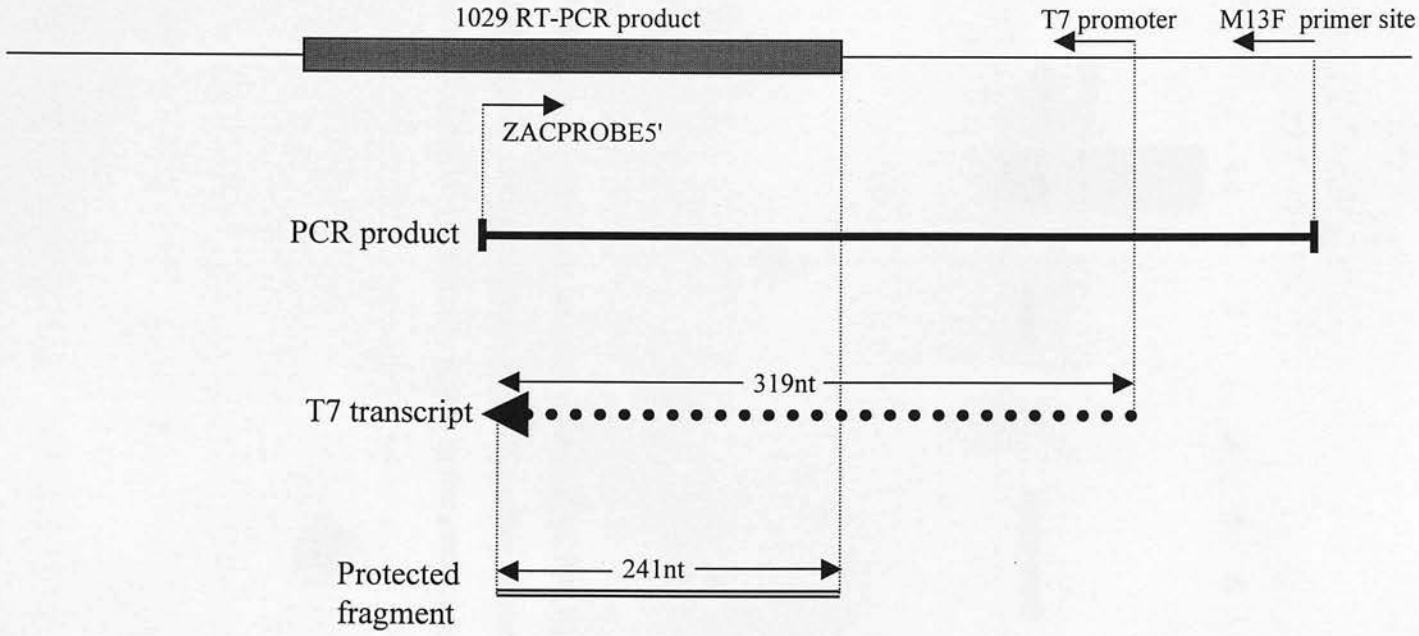


Figure 61. Generation of an antisense probe for ZAC. The RT-PCR product 1029 was cloned into the pGEM-T easy vector, then a PCR product generated from the clone using primers M13F and ZACPROBE5'. Transcription from the T7 promoter site yields an antisense probe of 319nt. Following hybridisation and digestion with ribonuclease, the expected protected fragment is 241nt.

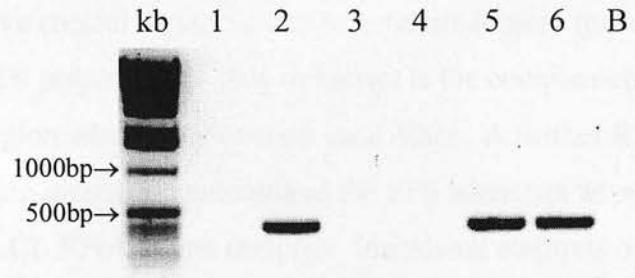


Figure 62. PCR from M13F to an internal primer ZACPROBE5', to determine the orientation of the insert. 1-6 are clones, B is a negative control PCR containing no template. Clones 2, 5 and 6 contain the insert in the correct orientation. See text for more details.

5.5.4.3 RPA experiments

Several pilot experiments were set up as recommended in the Ambion protocol, to try to optimise the amount of RNA and concentration of enzyme to use in the assay. However, despite generating a full-length probe in each transcription reaction, no protected fragment was observed after hybridisation and digestion. The sample RNA used in these pilot experiments was derived from a lymphoblastoid cell line. Because of these results, it was suspected therefore that the *ZAC* transcript might be too rare to detect by this technique. To ensure that the RPA method was working, the only positive control available was to generate a sense transcript from the original clone, using SP6 polymerase. This transcript is the complement of the T7 antisense probe in the region where they overlap each other. A further RPA experiment was performed, using increasing amounts of the SP6 transcript as positive control samples, and LCL RNA as test samples. Increasing amounts of SP6 transcript were used, ranging from 0.6pg to 600pg. These were chosen to represent the approximate equivalent amounts of *ZAC* transcript that would be expected in a 10μg sample of total RNA, assuming abundances of *ZAC* in the range between 1 in 100000 mRNA molecules (represented by 0.6pg SP6 transcript) and and 1/100 mRNA molecules (represented by 600pg SP6 transcript), taking into account the shorter length of the SP6 transcript as compared to a mRNA transcript of average length. The result is shown in Figure 63. Despite the apparent premature termination of the probe, there is no protected fragment in any of the LCL test lanes, but a protected fragment is evident in the SP6 control lane with the highest amount of SP6 transcript, equivalent to an abundance of 1/100 mRNA molecules, and is weakly visible in the lane representing an abundance of 1/1000 mRNA molecules. This result suggests that the experimental method is working, but that the *ZAC* transcript may be too rare to allow it to be detected by this technique. Published literature that has commented on the rarity of *ZAC* mRNA includes Basyuk *et al.* (Basyuk *et al.* 2000). However, it should be possible to detect the SP6 transcript at lower amounts than 60pg, since this is equivalent to an abundance of 1/1000 mRNA molecules, which is relatively high. It is possible therefore that either the quantification of the SP6 transcript was incorrect, or that the RPA parameters used here were not optimal.

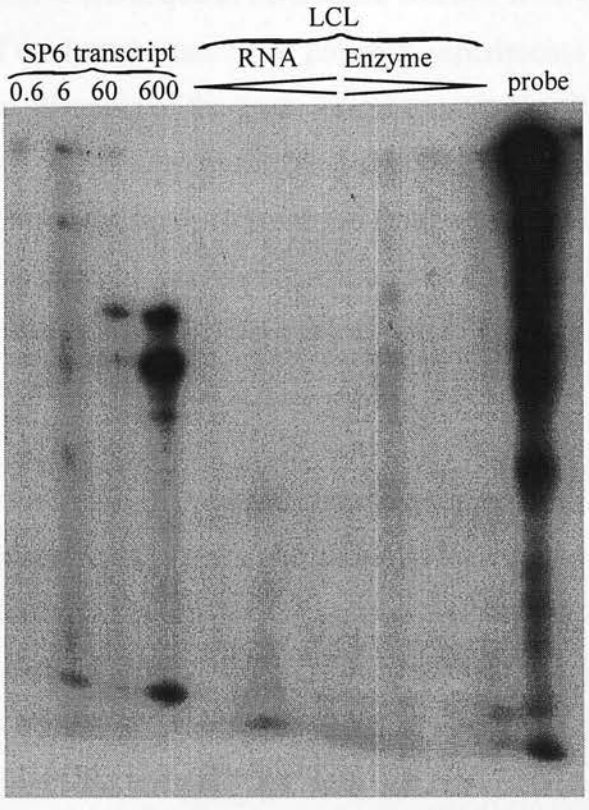


Figure 63. RPA experiment with SP6 transcript and LCL RNA as test samples. Increasing concentrations of 0.6pg to 600pg of SP6 transcript were used. For the lanes with LCL as the sample, increasing amounts of RNA (and constant amount of enzyme), and then decreasing amounts of enzyme (and constant amount of RNA) were used. The far right lane contains undigested probe.

5.5.5 Expression analysis by RT-PCR

As described, the rarity of the *ZAC* transcript may have precluded its detection by RPA. As an alternative technique to investigate whether the hypermethylation observed in two of the lymphomas in the previous experiments may be associated with transcriptional silencing of the gene, a semi-quantitative RT-PCR approach was chosen. This technique relies on quantifying the resulting RT-PCR product of the *ZAC* gene, in relation to the product generated from a typical housekeeping gene, such as *GAPDH*. Expressing the relative amounts of *ZAC* transcript as a proportion of *GAPDH* eliminates the variation between samples of the quality or amount of template in the reaction.

Total RNA was extracted from the same lymphoma samples used for the DNA analysis. In preliminary experiments, the timing of the exponential phase was determined for both the *GAPDH* RT-PCR and the *ZAC* RT-PCR (using primers 1029-RT5' and 1029-RT3' which were used previously in Chapter 4). A series of duplicate PCR reactions were set up and cycled for a range of cycle numbers. The templates for the RT-PCRs were cDNAs from two fetal RNA samples and two non-hypermethylated lymphomas. The results of this experiment are shown in Figure 64. It is apparent that whilst a product for *GAPDH* appears at 20 cycles in the fetal samples, it is not amplified visibly from the lymphoma samples until much later, at 26 cycles. This may indicate poor quality of the NHL RNAs, since *GAPDH* is assumed to be present at comparable levels in most cells. Similarly, for the *ZAC* RT-PCR, amplification from the fetal samples is visible at 30 cycles, but is only very weak in the lymphoma samples at 36 cycles with no visible product before this point. As with the RNase protection assay, this technique indicates the probable rarity of the *ZAC* transcript. However, because of the difference in the success of amplification of *GAPDH* between the fetal samples and the lymphomas, it is apparent that the quality of these RNA or cDNA samples is poor. Indeed, the lymphoma tissue samples had been stored for a lengthy period of time at -70°C before these experiments were carried out, and it is likely, as judged by these results, that some degradation had occurred. The difficulties in amplifying even a widely-

expressed housekeeping gene such as *GAPDH* suggest that a quantitative assessment of the relative levels of *ZAC* and *GAPDH* in these samples by this technique may not be possible using this archival material.

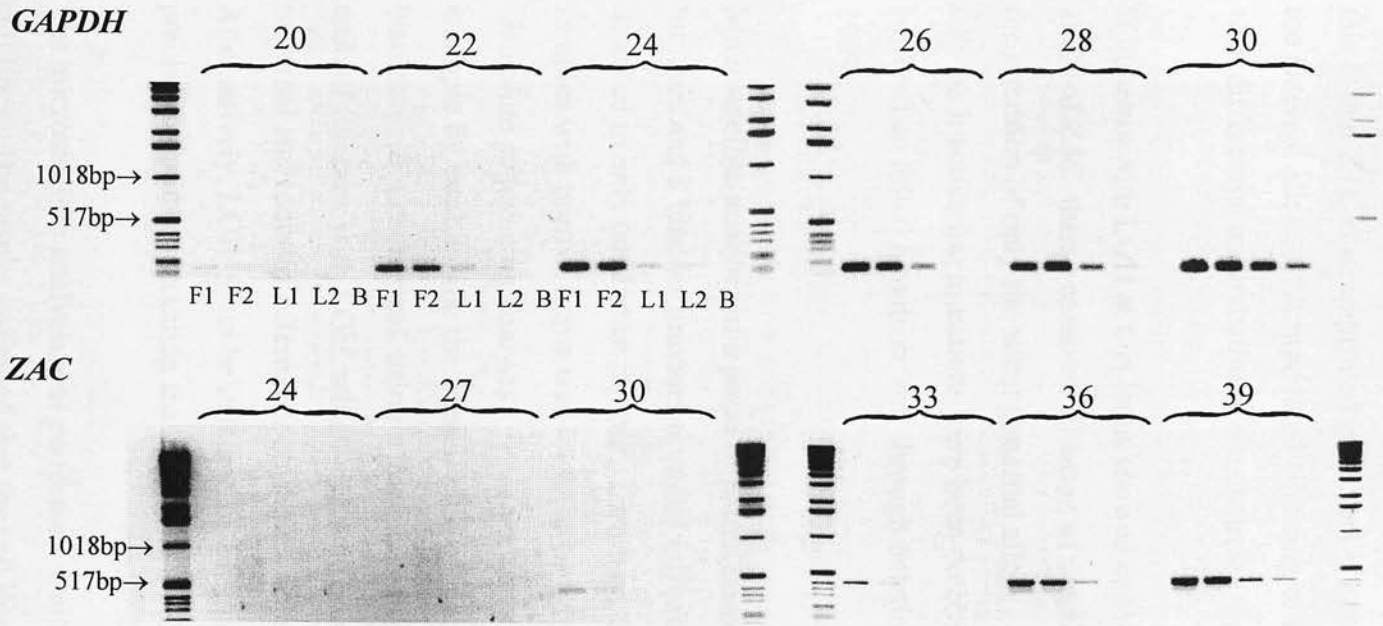


Figure 64. PCR series to determine the exponential phase for *GAPDH* and *ZAC* 1029 RT-PCRs, in preparation for semi-quantitative RT-PCR analysis. F1 and F2 are cDNAs of fetal muscle and fetal gut, respectively. L1 and L2 are cDNAs of lymphoma samples that were not hypermethylated at NV149. B is a negative control PCR containing no template. The order of these samples is the same on all gels. The numbers above the brackets represent the number of cycles. The marker is the 1kb ladder with band sizes as indicated on the left.

5.6 DISCUSSION

In this study, the initial hypothesis being tested predicted that the active paternal allele of *ZAC* will be preferentially lost in tumours carrying a deletion. The basis of this is that *ZAC* is an imprinted gene, and is known to be expressed exclusively from the paternal allele, with functional silencing of the maternal allele being associated with differential methylation of the upstream CpG island NV149.

If tumours with LOH at this locus showed consistent preferential loss of the active allele of *ZAC*, then a complete ablation of function would presumably result, due to the retention of only the silent maternal allele. This would offer an explanation as to why no inactivating mutations have been detected in *ZAC* in tumours which have suffered an initial Knudson 'hit' through deletion of this region (Bilanges *et al.* 1999).

Microsatellite analysis of a panel of paired normal and tumour DNAs from 11 breast tumours and 8 bladder tumours revealed infrequent LOH. In fact, possible LOH was detected in only one breast tumour. However, the contamination of these tumour samples with normal tissue was found to be a significant source of distortion in the bisulphite sequencing analysis. There are ways of overcoming this problem, for example by establishing the normal ratio of maternal to paternal alleles on the bisulphite sequencing gel, using a phosphorimager, and then measuring the altered ratio of tumours with LOH, which might be predicted to show a predominantly maternal and reduced paternal contribution as compared to the normal. Alternatively, LOH could be detected by FISH techniques, which could be more precise if a probe containing the *ZAC* gene itself was used.

The microsatellite analysis was performed similarly on a panel of six breast cancer cell lines. The results indicated that the cell line ADR-MCF7 carries a deletion of the *ZAC*/NV149 region, but the retained allele is predominantly unmethylated and therefore presumably the paternal (active) copy, rather than the maternal (inactive) copy. There is however some evidence of a methylated allele by the presence of a

faint lower 'A' band, so this may be representative of secondary methylation. LOH in this case would not implicate *ZAC* as a tumour suppressor gene as the cell line predominantly carries the unmethylated and presumably still functional paternal copy of *ZAC*. There was no other significant LOH at the *ZAC*/NV149 locus in the panel of cell lines, although the bisulphite analysis did nonetheless show that the cell line T47D has a predominantly maternal-like methylation pattern. This suggests that hypermethylation had occurred, and could implicate *ZAC* as a tumour suppressor gene, because both maternal and paternal chromosomes carry a maternal methylation pattern which would be predicted to silence the gene.

The panel of 21 non-Hodgkin's lymphomas also revealed very little LOH at the markers tested. However, the bisulphite sequencing, which was anticipated to show biparental methylation in all tumours (since both parental alleles are present) produced an unexpected result. While most of the lymphomas did show the predicted biparental methylation pattern, two were found to have only the maternal methylated pattern at NV149. As with the breast cancer cell line T47D, this result is suggestive of hypermethylation at NV149, which might result in silencing of both alleles, and thus a complete ablation of function of *ZAC* in these tumours.

This study was designed to test the hypothesis that imprinting of *ZAC* acts as a 'first hit', inactivating the maternal allele, and when accompanied by a 'second hit' deletion of the remaining active paternal allele, the tumour suppressor function of *ZAC* is eliminated, thus paving the way for tumour development. In fact, inactivation by this mechanism has not been found in the tumours analysed here, but there is evidence of hypermethylation as an alternative, epigenetic mechanism of *ZAC* inactivation in some lymphomas. Indeed, hypermethylation of tumour suppressor genes has been reported for several other genes in a variety of different cancers, for example *CDKN2A* ($p16^{INK4a}$) in breast, prostate, renal and colon cancers (Herman *et al.* 1995), *VHL* in renal cancer (Herman *et al.* 1994), *CDH1* (*E-cadherin*) in breast and prostate cancers (Graff *et al.* 1995), and *BRCA1* in sporadic breast and ovarian cancers (Esteller *et al.* 2000). In addition, the biallelic hypermethylation of normally imprinted tumour suppressor genes has also been reported: the *TP73* gene

has been found to be hypermethylated in lymphomas (Kawano *et al.* 1999; Corn *et al.* 1999), and *H19* is subject to hypermethylation in some Wilms' tumours and hepatoblastomas (Fukuzawa *et al.* 1999). Hypermethylation at NV149 could, similarly, be a mechanism of inactivation of *ZAC*, resulting in a loss of its anti-proliferative, anti-apoptotic function.

To support this idea, attempts were made to analyse *ZAC* expression in the lymphomas. However, the rarity of the *ZAC* transcript prevented quantification of this loss by RNase protection assay, since the transcript was undetectable even in normal LCL RNA. Furthermore, it became apparent that the quality of the RNA samples was insufficient to allow a valid quantification by RT-PCR.

To determine the role of hypermethylation as a mechanism of *ZAC* inactivation in lymphoma, further efforts will be required involving the analysis of a further, larger sample of lymphomas. An antibody capable of detecting ZAC protein would also be very useful, for assessing the levels of protein in tumours with LOH or hypermethylation compared to controls.

Future work could focus on establishing the extent to which this phenomenon may be involved in tumorigenesis, as well as the actual mechanisms by which it occurs. In the panel of tumours in this study, an extensive mutation screening effort may identify inactivating point mutations in those samples which had neither LOH nor hypermethylation, because if *ZAC* is acting as a tumour suppressor gene in these samples, then they might be predicted to have acquired a second 'hit' in the form of a mutation.

Extending the analysis to a wider range of tumour types would also be interesting, particularly in those tumour types which, like non-Hodgkins' lymphoma, have a high frequency of LOH at this chromosomal region. Ultimately, a transgenic *ZAC* knockout mouse would be of great interest to allow an assessment of the effects of a complete loss of the *ZAC* gene product.

REFERENCES

Abdollahi A, Godwin AK, Miller PD, Getts LA, Schultz DC, Taguchi T, Testa JR, Hamilton TC (1997a) Identification of a gene containing zinc-finger motifs based on lost expression in malignantly transformed rat ovarian surface epithelial cells. *Cancer Res* **57**:2029-34

Abdollahi A, Roberts D, Godwin AK, Schultz DC, Sonoda G, Testa JR, Hamilton TC (1997b) Identification of a zinc-finger gene at 6q25: a chromosomal region implicated in development of many solid tumors. *Oncogene* **14**:1973-9

Abe T, Makino N, Furukawa T, Ouyang H, Kimura M, Yatsuoka T, Yokoyama T, Inoue H, Fukushige S, Hoshi M, Hayashi Y, Sunamura M, Kobari M, Matsuno S, Horii A (1999) Identification of three commonly deleted regions on chromosome arm 6q in human pancreatic cancer. *Genes Chromosomes Cancer* **25**:60-4

Arthur EI, Zlotogora J, Lerer I, Dagan J, Marks K, Abeliovich D (1997) Transient neonatal diabetes mellitus in a child with invdup(6)(q22q23) of paternal origin. *Eur J Hum Genet* **5**:417-9

Barlow DP (1993) Methylation and imprinting: from host defense to gene regulation? *Science* **260**:309-10

Barlow DP (1995) Gametic imprinting in mammals. *Science* **270**:1610-3

Bartolomei MS, Webber AL, Brunkow ME, Tilghman SM (1993) Epigenetic mechanisms underlying the imprinting of the mouse H19 gene. *Genes Dev* **7**:1663-73

Barton SC, Surani MA, Norris ML (1984) Role of paternal and maternal genomes in mouse development. *Nature* 311:374-6

Basyuk E, Bertrand E, Journot L (2000) Alkaline fixation drastically improves the signal of in situ hybridization. *Nucleic Acids Res* **28**:E46

Bell AC, Felsenfeld G (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**:482-5

Bell DW, Jhanwar SC, Testa JR (1997) Multiple regions of allelic loss from chromosome arm 6q in malignant mesothelioma. *Cancer Res* **57**:4057-62

Bernues M, Casadevall C, Caballin MR, Miro R, Ejarque MJ, Chechile G, Gelabert A, Egozcue J (1999) Study of allelic losses on 3p, 6q, and 17p in human urothelial cancer. *Cancer Genet Cytogenet* **112**:42-5

Bielinska B, Blaydes SM, Buiting K, Yang T, Krajewska-Walasek M, Horsthemke B, Brannan CI (2000) De novo deletions of SNRPN exon 1 in early human and mouse embryos result in a paternal to maternal imprint switch. *Nat Genet* **25**:74-8

Bilanges B, Varrault A, Basyuk E, Rodriguez C, Mazumdar A, Pantaloni C, Bockaert J, Theillet C, Spengler D, Journot L (1999) Loss of expression of the

candidate tumor suppressor gene ZAC in breast cancer cell lines and primary tumors. *Oncogene* **18**:3979-88

Bird A (1992) The essentials of DNA methylation. *Cell* **70**:5-8

Blagitko N, Schulz U, Schinzel AA, Ropers HH, Kalscheuer VM (1999) gamma2-COP, a novel imprinted gene on chromosome 7q32, defines a new imprinting cluster in the human genome. *Hum Mol Genet* **8**:2387-96

Blagitko N, Mergenthaler S, Schulz U, Wollmann HA, Craigen W, Eggermann T, Ropers HH, Kalscheuer VM (2000) Human GRB10 is imprinted and expressed from the paternal and maternal allele in a highly tissue- and isoform-specific fashion. *Hum Mol Genet* **9**:1587-95

Bloomfield CD, Arthur DC, Frizzera G, Levine EG, Peterson BA, Gajl-Peczalska KJ (1983) Nonrandom chromosome abnormalities in lymphoma. *Cancer Res* **43**:2975-84

Boccaccio I, Glatt-Deeley H, Watrin F, Roeckel N, Lalande M, Muscatelli F (1999) The human MAGEL2 gene and its mouse homologue are paternally expressed and mapped to the Prader-Willi region. *Hum Mol Genet* **8**:2497-505

Brandeis M, Kafri T, Ariel M, Chaillet JR, McCarrey J, Razin A, Cedar H (1993) The ontogeny of allele-specific methylation associated with imprinted genes in the mouse. *Embo J* **12**:3669-77

Bravard A, Sabatier L, Hoffschir F, Ricoul M, Luccioni C, Dutrillaux B (1992) SOD2: a new type of tumor-suppressor gene? *Int J Cancer* **51**:476-80

Brodeur GM, Green AA, Hayes FA, Williams KJ, Williams DL, Tsiatis AA (1981) Cytogenetic features of human neuroblastomas and cell lines. *Cancer Res* **41**:4678-86

Brodeur GM, Seeger RC, Schwab M, Varmus HE, Bishop JM (1984) Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* **224**:1121-4

Buckley MF, Loveland KA, McKinstry WJ, Garson OM, Goding JW (1990) Plasma cell membrane glycoprotein PC-1. cDNA cloning of the human molecule, amino acid sequence, and chromosomal location. *J Biol Chem* **265**:17506-11

Buiting K, Saitoh S, Gross S, Dittrich B, Schwartz S, Nicholls RD, Horsthemke B (1995) Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. *Nat Genet* **9**:395-400

Cai YC, Yang GY, Nie Y, Wang LD, Zhao X, Song YL, Seril DN, Liao J, Xing EP, Yang CS (2000) Molecular alterations of p73 in human esophageal squamous cell carcinomas: loss of heterozygosity occurs frequently; loss of imprinting and

elevation of p73 expression may be related to defective p53. *Carcinogenesis* **21**:683-9

Campbell R, Gosden CM, Bonthron DT (1994) Parental origin of transcription from the human GNAS1 gene. *J Med Genet* **31**:607-14

Caron H, van Sluis P, van Hoeve M, de Kraker J, Bras J, Slater R, Mannens M, Voute PA, Westerveld A, Versteeg R (1993) Allelic loss of chromosome 1p36 in neuroblastoma is of preferential maternal origin and correlates with N-myc amplification. *Nat Genet* **4**:187-90

Caron H, Peter M, van Sluis P, Speleman F, de Kraker J, Laureys G, Michon J, Brugieres L, Voute PA, Westerveld A, et al. (1995) Evidence for two tumour suppressor loci on chromosomal bands 1p35-36 involved in neuroblastoma: one probably imprinted, another associated with N-myc amplification. *Hum Mol Genet* **4**:535-9

Caspary T, Cleary MA, Baker CC, Guan XJ, Tilghman SM (1998) Multiple mechanisms regulate imprinting of the mouse distal chromosome 7 gene cluster. *Mol Cell Biol* **18**:3466-74

Caspary T, Cleary MA, Perlman EJ, Zhang P, Elledge SJ, Tilghman SM (1999) Oppositely imprinted genes p57(Kip2) and igf2 interact in a mouse model for Beckwith-Wiedemann syndrome. *Genes Dev* **13**:3115-24

Cattanach BM, Kirk M (1985) Differential activity of maternally and paternally derived chromosome regions in mice. *Nature* **315**:496-8

Cattanach BM, Barr JA, Beechey CV, Martin J, Noebels J, Jones J (1997) A candidate model for Angelman syndrome in the mouse. *Mamm Genome* **8**:472-8

Cave H, Polak M, Drunat S, Denamur E, Czernichow P (2000) Refinement of the 6q chromosomal region implicated in transient neonatal diabetes. *Diabetes* **49**:108-13

Chaudhuri S, Messing J (1994) Allele-specific parental imprinting of dzr1, a posttranscriptional regulator of zein accumulation. *Proc Natl Acad Sci U S A* **91**:4867-71

Chen CL, Ip SM, Cheng D, Wong LC, Ngan HY (2000a) Loss of imprinting of the IGF-II and H19 genes in epithelial ovarian cancer. *Clin Cancer Res* **6**:474-9

Chen CL, Ip SM, Cheng D, Wong LC, Ngan HY (2000b) P73 gene expression in ovarian cancer tissues and cell lines. *Clin Cancer Res* **6**:3910-5

Cheng JM, Hiemstra JL, Schneider SS, Naumova A, Cheung NK, Cohn SL, Diller L, Sapienza C, Brodeur GM (1993) Preferential amplification of the paternal allele of the N-myc gene in human neuroblastomas. *Nat Genet* **4**:191-4

Cheng NC, Van Roy N, Chan A, Beitsma M, Westerveld A, Speleman F, Versteeg R (1995) Deletion mapping in neuroblastoma cell lines suggests two distinct tumor suppressor genes in the 1p35-36 region, only one of which is associated with N-myc amplification. *Oncogene* **10**:291-7

Cher ML, Bova GS, Moore DH, Small EJ, Carroll PR, Pin SS, Epstein JI, Isaacs WB, Jensen RH (1996) Genetic alterations in untreated metastases and androgen-independent prostate cancer detected by comparative genomic hybridization and allelotyping. *Cancer Res* **56**:3091-102

Chi SG, Chang SG, Lee SJ, Lee CH, Kim JI, Park JH (1999) Elevated and biallelic expression of p73 is associated with progression of human bladder cancer. *Cancer Res* **59**:2791-3

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **162**:156-9

Chumakov I, Rigault P, Guillou S, Ougen P, Billaut A, Guasconi G, Gervy P, LeGall I, Soularue P, Grinas L, et al. (1992) Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature* **359**:380-7

Clark GM, McGuire WL (1988) Steroid receptors and other prognostic factors in primary breast cancer. *Semin Oncol* **15**:20-5

Clark SJ, Harrison J, Paul CL, Frommer M (1994) High sensitivity mapping of methylated cytosines. *Nucleic Acids Res* **22**:2990-7

Coffey JD, Jr., Womack NC (1967) Transient neonatal diabetes mellitus in half sisters. *Am J Dis Child* **113**:480-2

Colitti CV, Rodabaugh KJ, Welch WR, Berkowitz RS, Mok SC (1998) A novel 4 cM minimal deletion unit on chromosome 6q25.1-q25.2 associated with high grade invasive epithelial ovarian carcinomas. *Oncogene* **16**:555-9

Constancia M, Pickard B, Kelsey G, Reik W (1998) Imprinting mechanisms. *Genome Res* **8**:881-900

Cooke IE, Shelling AN, Le Meuth VG, Charnock ML, Ganesan TS (1996) Allele loss on chromosome arm 6q and fine mapping of the region at 6q27 in epithelial ovarian cancer. *Genes Chromosomes Cancer* **15**:223-33

Cooney KA, Wetzel JC, Consolino CM, Wojno KJ (1996) Identification and characterization of proximal 6q deletions in prostate cancer. *Cancer Res* **56**:4150-3

Corn PG, Kuerbitz SJ, van Noesel MM, Esteller M, Compitello N, Baylin SB, Herman JG (1999) Transcriptional silencing of the p73 gene in acute lymphoblastic leukemia and Burkitt's lymphoma is associated with 5' CpG island methylation. *Cancer Res* **59**:3352-6

Costello JF, Fruhwald MC, Smiraglia DJ, Rush LJ, Robertson GP, Gao X, Wright FA, Feramisco JD, Peltomaki P, Lang JC, Schuller DE, Yu L, Bloomfield CD, Caligiuri MA, Yates A, Nishikawa R, Su Huang H, Petrelli NJ, Zhang X, O'Dorisio MS, Held WA, Cavenee WK, Plass C (2000) Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet* **24**:132-8

Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, Reed PW, Gough SC, Jenkins SC, Palmer SM, et al. (1994) A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* **371**:130-6

Davies SJ, Hughes HE (1993) Imprinting in Albright's hereditary osteodystrophy. *J Med Genet* **30**:101-3

Davis TL, Trasler JM, Moss SB, Yang GJ, Bartolomei MS (1999) Acquisition of the H19 methylation imprint occurs differentially on the parental alleles during spermatogenesis. *Genomics* **58**:18-28

Davis TL, Yang GJ, McCarrey JR, Bartolomei MS (2000) The H19 methylation imprint is erased and re-established differentially on the parental alleles during male germ cell development. *Hum Mol Genet* **9**:2885-94

Dawson WD (1965) Fertility and size inheritance in a *Peromyscus* species cross. *Evolution* **19**:44-55

DeChiara TM, Robertson EJ, Efstratiadis A (1991) Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* **64**:849-59

Delepine M, Pociot F, Habita C, Hashimoto L, Froguel P, Rotter J, Cambon-Thomsen A, Deschamps I, Djoulah S, Weissenbach J, Nerup J, Lathrop M, Julier C (1997) Evidence of a non-MHC susceptibility locus in type I diabetes linked to HLA on chromosome 6. *Am J Hum Genet* **60**:174-87

Douc-Rasy S, Barrois M, Fogel S, Ahomadegbe JC, Stehelin D, Coll J, Riou G (1996) High incidence of loss of heterozygosity and abnormal imprinting of H19 and IGF2 genes in invasive cervical carcinomas. Uncoupling of H19 and IGF2 expression and biallelic hypomethylation of H19. *Oncogene* **12**:423-30

Drewell RA, Brenton JD, Ainscough JF, Barton SC, Hilton KJ, Arney KL, Dandolo L, Surani MA (2000) Deletion of a silencer element disrupts H19 imprinting independently of a DNA methylation epigenetic switch. *Development* **127**:3419-28

Dutrillaux B, Gerbault-Seureau M, Zafrani B (1990) Characterization of chromosomal anomalies in human breast cancer. A comparison of 30 paradiploid cases with few chromosome changes. *Cancer Genet Cytogenet* **49**:203-17

Eggerding FA, Schonberg SA, Chehab FF, Norton ME, Cox VA, Epstein CJ (1994) Uniparental isodisomy for paternal 7p and maternal 7q in a child with growth retardation. *Am J Hum Genet* **55**:253-65

Ejeskar K, Sjoberg RM, Kogner P, Martinsson T (1999) Variable expression and absence of mutations in p73 in primary neuroblastoma tumors argues against a role in neuroblastoma development. *Int J Mol Med* **3**:585-9

Ejeskar K, Abel F, Sjoberg R, Backstrom J, Kogner P, Martinsson T (2000) Fine mapping of the human preprocortistatin gene (CORT) to neuroblastoma consensus deletion region 1p36.3-p36.2, but absence of mutations in primary tumors. *Cytogenet Cell Genet* **89**:62-6

Ellmeier W, Aguzzi A, Kleiner E, Kurzbauer R, Weith A (1992) Mutually exclusive expression of a helix-loop-helix gene and N-myc in human neuroblastomas and in normal development. *EMBO J* **11**:2563-71

el-Naggar AK, Lai S, Tucker SA, Clayman GL, Goepfert H, Hong WK, Huff V (1999) Frequent loss of imprinting at the IGF2 and H19 genes in head and neck squamous carcinoma. *Oncogene* **18**:7063-9

Enari M, Sakahira H, Yokoyama H, Okawa K, Iwamatsu A, Nagata S (1998) A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. *Nature* **391**:43-50

Engemann S, Strodicke M, Paulsen M, Franck O, Reinhardt R, Lane N, Reik W, Walter J (2000) Sequence and functional comparison in the beckwith-wiedemann region: implications for a novel imprinting centre and extended imprinting. *Hum Mol Genet* **9**:2691-706

Enomoto H, Ozaki T, Takahashi E, Nomura N, Tabata S, Takahashi H, Ohnuma N, Tanabe M, Iwai J, Yoshida H, et al. (1994) Identification of human DAN gene, mapping to the putative neuroblastoma tumor suppressor locus. *Oncogene* **9**:2785-91

Esteller M, Silva JM, Dominguez G, Bonilla F, Matias-Guiu X, Lerma E, Bussaglia E, Prat J, Harkes IC, Repasky EA, Gabrielson E, Schutte M, Baylin SB, Herman JG (2000) Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *J Natl Cancer Inst* **92**:564-9

Feil R, Boyano MD, Allen ND, Kelsey G (1997) Parental chromosome-specific chromatin conformation in the imprinted U2af1-rs1 gene in the mouse. *J Biol Chem* **272**:20893-900

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* **132**:6-13

Ferguson AW, Milner RD (1970) Transient neonatal diabetes mellitus in sibs. *Arch Dis Child* **45**:80-3

Ferguson-Smith AC, Sasaki H, Cattanach BM, Surani MA (1993) Parental-origin-specific epigenetic modification of the mouse H19 gene. *Nature* **362**:751-5

Ferguson-Smith AC (2000) Genetic imprinting: silencing elements have their say. *Curr Biol* **10**:R872-5

Filipsson K, Sundler F, Ahren B (1999) PACAP is an islet neuropeptide which contributes to glucose-stimulated insulin secretion. *Biochem Biophys Res Commun* **256**:664-7

Fong CT, Dracopoli NC, White PS, Merrill PT, Griffith RC, Housman DE, Brodeur GM (1989) Loss of heterozygosity for the short arm of chromosome 1 in human neuroblastomas: correlation with N-myc amplification. *Proc Natl Acad Sci U S A* **86**:3753-7

Foulkes WD, Ragoussis J, Stamp GW, Allan GJ, Trowsdale J (1993) Frequent loss of heterozygosity on chromosome 6 in human ovarian carcinoma. *Br J Cancer* **67**:551-9

Fujii H, Zhou W, Gabrielson E (1996) Detection of frequent allelic loss of 6q23-q25.2 in microdissected human breast cancer tissues. *Genes Chromosomes Cancer* **16**:35-9

Fukuzawa R, Umezawa A, Ochi K, Urano F, Ikeda H, Hata J (1999) High frequency of inactivation of the imprinted H19 gene in sporadic hepatoblastoma. *Int J Cancer* **82**:490-7

Gaidano G, Hauptschein RS, Parsa NZ, Offit K, Rao PH, Lenoir G, Knowles DM, Chaganti RS, Dalla-Favera R (1992) Deletions involving two distinct regions of 6q in B-cell non-Hodgkin lymphoma. *Blood* **80**:1781-7

Gardner RJ, Mungall AJ, Dunham I, Barber JC, Shield JP, Temple IK, Robinson DO (1999) Localisation of a gene for transient neonatal diabetes mellitus to an 18.72 cR3000 (approximately 5.4 Mb) interval on chromosome 6q. *J Med Genet* **36**:192-6

Gardner RJ, Mackay DJ, Mungall AJ, Polychronakos C, Siebert R, Shield JP, Temple IK, Robinson DO (2000) An imprinted locus associated with transient neonatal diabetes mellitus. *Hum Mol Genet* **9**:589-96

Goldberg EK, Glendening JM, Karanjawala Z, Sridhar A, Walker GJ, Hayward NK, Rice AJ, Kurera D, Tebha Y, Fountain JW (2000) Localization of multiple melanoma tumor-suppressor genes on chromosome 11 by use of homozygosity mapping-of-deletions analysis. *Am J Hum Genet* **67**:417-31

Graff JR, Herman JG, Lapidus RG, Chopra H, Xu R, Jarrard DF, Isaacs WB, Pitha PM, Davidson NE, Baylin SB (1995) E-cadherin expression is silenced by DNA hypermethylation in human breast and prostate carcinomas. *Cancer Res* **55**:5195-9

Grandjean V, Smith J, Schofield PN, Ferguson-Smith AC (2000) Increased IGF-II protein affects p57kip2 expression in vivo and in vitro: implications for Beckwith-Wiedemann syndrome. *Proc Natl Acad Sci U S A* **97**:5279-84

Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillett D, Muselet D, Prud'Homme JF, Dib C, Auffray C, Morissette J, Weissenbach J, Goodfellow PN (1996) A radiation hybrid map of the human genome. *Hum Mol Genet* **5**:339-46

Haig D (1994) Refusing the ovarian time bomb. *Trends Genet* **10**:346-7; discussion 348-9

Halenbeck R, MacDonald H, Roulston A, Chen TT, Conroy L, Williams LT (1998) CPAN, a human nuclease regulated by the caspase-sensitive inhibitor DFF45. *Curr Biol* **8**:537-40

Han S, Semba S, Abe T, Makino N, Furukawa T, Fukushige S, Takahashi H, Sakurada A, Sato M, Shiiba K, Matsuno S, Nimura Y, Nakagawara A, Horii A (1999) Infrequent somatic mutations of the p73 gene in various human cancers. *Eur J Surg Oncol* **25**:194-8

Hankins GR, De Souza AT, Bentley RC, Patel MR, Marks JR, Iglehart JD, Jirtle RL (1996) M6P/IGF2 receptor: a candidate breast tumor suppressor gene. *Oncogene* **12**:2003-9

Hannula K, Lipsanen-Nyman M, Kontiokari T, Kere J (2001) A Narrow Segment of Maternal Uniparental Disomy of Chromosome 7q31-qter in Silver-Russell Syndrome Delimits a Candidate Gene Region. *Am J Hum Genet* **68**:247-253

Hao Y, Crenshaw T, Moulton T, Newcomb E, Tycko B (1993) Tumour-suppressor activity of H19 RNA. *Nature* **365**:764-7

Hark AT, Tilghman SM (1998) Chromatin conformation of the H19 epigenetic mark. *Hum Mol Genet* **7**:1979-85

Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405**:486-9

Hatada I, Mukai T (1995) Genomic imprinting of p57KIP2, a cyclin-dependent kinase inhibitor, in mouse. *Nat Genet* **11**:204-6

Hauptschein RS, Gamberi B, Rao PH, Frigeri F, Scotto L, Venkatraj VS, Gaidano G, Rutner T, Edwards YH, Chaganti RS, Dalla-Favera R (1998) Cloning and mapping of human chromosome 6q26-q27 deleted in B-cell non-Hodgkin lymphoma and multiple tumor types. *Genomics* **50**:170-86

Hayward BE, Kamiya M, Strain L, Moran V, Campbell R, Hayashizaki Y, Bonthron DT (1998a) The human GNAS1 gene is imprinted and encodes distinct paternally and biallelically expressed G proteins. *Proc Natl Acad Sci U S A* **95**:10038-43

Hayward BE, Moran V, Strain L, Bonthron DT (1998b) Bidirectional imprinting of a single gene: GNAS1 encodes maternally, paternally, and biallelically derived proteins. *Proc Natl Acad Sci U S A* **95**:15475-80

Hayward BE, Bonthron DT (2000) An imprinted antisense transcript at the human GNAS1 locus. *Hum Mol Genet* **9**:835-41

Henry I, Bonaiti-Pellie C, Chehensse V, Beldjord C, Schwartz C, Utermann G, Junien C (1991) Uniparental paternal disomy in a genetic cancer-predisposing syndrome. *Nature* **351**:665-7

Herman JG, Latif F, Weng Y, Lerman MI, Zbar B, Liu S, Samid D, Duan DS, Gnarra JR, Linehan WM, et al. (1994) Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc Natl Acad Sci U S A* **91**:9700-4

Herman JG, Merlo A, Mao L, Lapidus RG, Issa JP, Davidson NE, Sidransky D, Baylin SB (1995) Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. *Cancer Res* **55**:4525-30

Hoglund P, Holmberg C, de la Chapelle A, Kere J (1994) Paternal isodisomy for chromosome 7 is compatible with normal growth and development in a patient with congenital chloride diarrhea. *Am J Hum Genet* **55**:747-52

Horike S, Mitsuya K, Meguro M, Kotobuki N, Kashiwagi A, Notsu T, Schulz TC, Shirayoshi Y, Oshimura M (2000) Targeted disruption of the human LIT1 locus defines a putative imprinting control element playing an essential role in beckwith-wiedemann syndrome. *Hum Mol Genet* **9**:2075-83

Hu JF, Ulaner GA, Oruganti H, Ivaturi RD, Balagura KA, Pham J, Vu TH, Hoffman AR (2000) Allelic expression of the putative tumor suppressor gene p73 in human fetal tissues and tumor specimens. *Biochim Biophys Acta* **1491**:49-56

Huang LW, Garrett AP, Muto MG, Colitti CV, Bell DA, Welch WR, Berkowitz RS, Mok SC (2000) Identification of a novel 9 cM deletion unit on chromosome 6q23-24 in papillary serous carcinoma of the peritoneum. *Hum Pathol* **31**:367-73

Hurst LD, McVean GT (1998) Do we understand the evolution of genomic imprinting? *Curr Opin Genet Dev* **8**:701-8

Ichimiya S, Nimura Y, Kageyama H, Takada N, Sunahara M, Shishikura T, Nakamura Y, Sakiyama S, Seki N, Ohira M, Kaneko Y, McKeon F, Caput D, Nakagawara A (1999) p73 at chromosome 1p36.3 is lost in advanced stage neuroblastoma but its mutation is infrequent. *Oncogene* **18**:1061-6

Ish-Horowicz D, Burke JF (1981) Rapid and efficient cosmid cloning. *Nucleic Acids Res* **9**:2989-98

Jay P, Rougeulle C, Massacrier A, Moncla A, Mattei MG, Malzac P, Roeckel N, Taviaux S, Lefranc JL, Cau P, Berta P, Lalande M, Muscatelli F (1997) The human necdin gene, NDN, is maternally imprinted and located in the Prader-Willi syndrome chromosomal region. *Nat Genet* **17**:357-61

Jones PL, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J, Wolffe AP (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* **19**:187-91

Jong MT, Carey AH, Caldwell KA, Lau MH, Handel MA, Driscoll DJ, Stewart CL, Rinchik EM, Nicholls RD (1999a) Imprinting of a RING zinc-finger encoding gene in the mouse chromosome region homologous to the Prader-Willi syndrome genetic region. *Hum Mol Genet* **8**:795-803

Jong MT, Gray TA, Ji Y, Glenn CC, Saitoh S, Driscoll DJ, Nicholls RD (1999b) A novel imprinted gene, encoding a RING zinc-finger protein, and overlapping antisense transcript in the Prader-Willi syndrome critical region. *Hum Mol Genet* **8**:783-93

Jost CA, Marin MC, Kaelin WG, Jr. (1997) p73 is a simian p53-related protein that can induce apoptosis. *Nature* **389**:191-4

Joyce CA, Sharp A, Walker JM, Bulliman H, Temple IK (1999) Duplication of 7p12.1-p13, including GRB10 and IGFBP1, in a mother and daughter with features of Silver-Russell syndrome. *Hum Genet* **105**:273-80

Joyce JA, Lam WK, Catchpoole DJ, Jenks P, Reik W, Maher ER, Schofield PN (1997) Imprinting of IGF2 and H19: lack of reciprocity in sporadic Beckwith-Wiedemann syndrome. *Hum Mol Genet* **6**:1543-8

Judson H, van Roy N, Strain L, Vandesompele J, Van Gele M, Speleman F, Bonthron DT (2000) Structure and mutation analysis of the gene encoding DNA fragmentation factor 40 (caspase-activated nuclease), a candidate neuroblastoma tumour suppressor gene. *Hum Genet* **106**:406-13

Kafri T, Ariel M, Brandeis M, Shemer R, Urven L, McCarrey J, Cedar H, Razin A (1992) Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes Dev* **6**:705-14

Kaghad M, Bonnet H, Yang A, Creancier L, Biscan JC, Valent A, Minty A, Chalon P, Lelias JM, Dumont X, Ferrara P, McKeon F, Caput D (1997) Monoallelically expressed gene related to p53 at 1p36, a region frequently deleted in neuroblastoma and other human cancers. *Cell* **90**:809-19

Kajii T, Ohama K (1977) Androgenetic origin of hydatidiform mole. *Nature* **268**:633-4

Kalscheuer VM, Mariman EC, Schepens MT, Rehder H, Ropers HH (1993) The insulin-like growth factor type-2 receptor gene is imprinted in the mouse but not in humans. *Nat Genet* **5**:74-8

Kamiya M, Judson H, Okazaki Y, Kusakabe M, Muramatsu M, Takada S, Takagi N, Arima T, Wake N, Kamimura K, Satomura K, Hermann R, Bonthron DT, Hayashizaki Y (2000) The cell cycle control gene ZAC/PLAGL1 is imprinted--a strong candidate gene for transient neonatal diabetes. *Hum Mol Genet* **9**:453-60

Kanduri C, Pant V, Loukinov D, Pugacheva E, Qi CF, Wolffe A, Ohlsson R, Lobanenkov VV (2000) Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr Biol* **10**:853-856

Kaneko-Ishino T, Kuroiwa Y, Miyoshi N, Kohda T, Suzuki R, Yokoyama M, Viville S, Barton SC, Ishino F, Surani MA (1995) Peg1/Mest imprinted gene on chromosome 6 identified by cDNA subtraction hybridization. *Nat Genet* **11**:52-9

Kang MJ, Park BJ, Byun DS, Park JI, Kim HJ, Park JH, Chi SG (2000) Loss of imprinting and elevated expression of wild-type p73 in human gastric adenocarcinoma. *Clin Cancer Res* **6**:1767-71

Kas K, Voz ML, Hensen K, Meyen E, Van de Ven WJ (1998) Transcriptional activation capacity of the novel PLAG family of zinc finger proteins. *J Biol Chem* **273**:23026-32

Kawano S, Miller CW, Gombart AF, Bartram CR, Matsuo Y, Asou H, Sakashita A, Said J, Tatsumi E, Koeffler HP (1999) Loss of p73 gene expression in leukemias/lymphomas due to hypermethylation. *Blood* **94**:1113-20

Kelsell DP, Rooke L, Warne D, Bouzyk M, Cullin L, Cox S, West L, Povey S, Spurr NK (1995) Development of a panel of monochromosomal somatic cell hybrids for rapid gene mapping. *Ann Hum Genet* **59**:233-41

Kelsey G, Reik W (1998) Analysis and identification of imprinted genes. *Methods* **14**:211-34

Kerjean A, Dupont JM, Vasseur C, Le Tessier D, Cuisset L, Paldi A, Jouannet P, Jeanpierre M (2000) Establishment of the paternal methylation imprint of the human H19 and MEST/PEG1 genes during spermatogenesis. *Hum Mol Genet* **9**:2183-7

Kim HT, Choi BH, Niikawa N, Lee TS, Chang SI (1998) Frequent loss of imprinting of the H19 and IGF-II genes in ovarian tumors. *Am J Med Genet* **80**:391-5

Kim KS, Lee YI (1997) Biallelic expression of the H19 and IGF2 genes in hepatocellular carcinoma. *Cancer Lett* **119**:143-8

Kinouchi Y, Hiwatashi N, Higashioka S, Nagashima F, Chida M, Toyota T (1996) Relaxation of imprinting of the insulin-like growth factor II gene in colorectal cancer. *Cancer Lett* **107**:105-8

Kirchhoff M, Rose H, Petersen BL, Maahr J, Gerdes T, Lundsteen C, Bryndorf T, Kryger-Baggesen N, Christensen L, Engelholm SA, Philip J (1999) Comparative genomic hybridization reveals a recurrent pattern of chromosomal aberrations in severe dysplasia/carcinoma in situ of the cervix and in advanced-stage cervical carcinoma. *Genes Chromosomes Cancer* **24**:144-50

Kishino T, Lalande M, Wagstaff J (1997) UBE3A/E6-AP mutations cause Angelman syndrome. *Nat Genet* **15**:70-3

Kitsberg D, Selig S, Brandeis M, Simon I, Keshet I, Driscoll DJ, Nicholls RD, Cedar H (1993) Allele-specific replication timing of imprinted gene regions. *Nature* **364**:459-63

Knoll JH, Nicholls RD, Magenis RE, Graham JM, Jr., Lalande M, Latt SA (1989) Angelman and Prader-Willi syndromes share a common chromosome 15 deletion but differ in parental origin of the deletion. *Am J Med Genet* **32**:285-90

Knoll JH, Glatt KA, Nicholls RD, Malcolm S, Lalande M (1991) Chromosome 15 uniparental disomy is not frequent in Angelman syndrome. *Am J Hum Genet* **48**:16-21

Knudson AG, Jr. (1971) Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* **68**:820-3

Knudson AG, Jr. (1986) Genetics of human cancer. *Annu Rev Genet* **20**:231-51

Knutsen T (1998) Cytogenetic changes in the progression of lymphoma. *Leuk Lymphoma* **31**:1-19

Kobayashi S, Kohda T, Miyoshi N, Kuroiwa Y, Aisaka K, Tsutsumi O, Kaneko-Ishino T, Ishino F (1997) Human PEG1/MEST, an imprinted gene on chromosome 7. *Hum Mol Genet* **6**:781-6

Kondo M, Suzuki H, Ueda R, Osada H, Takagi K, Takahashi T (1995) Frequent loss of imprinting of the H19 gene is often associated with its overexpression in human lung cancers. *Oncogene* **10**:1193-8

Kong FM, Anscher MS, Washington MK, Killian JK, Jirtle RL (2000) M6P/IGF2R is mutated in squamous cell carcinoma of the lung. *Oncogene* **19**:1572-8

Kosaki K, Kosaki R, Craigen WJ, Matsuo N (2000) Isoform-specific imprinting of the human PEG1/MEST gene. *Am J Hum Genet* **66**:309-12

Kotzot D, Schmitt S, Bernasconi F, Robinson WP, Lurie IW, Ilyina H, Mehes K, Hamel BC, Otten BJ, Hergersberg M, et al. (1995) Uniparental disomy 7 in Silver-Russell syndrome and primordial growth retardation. *Hum Mol Genet* **4**:583-7

Kovalev S, Marchenko N, Swendeman S, LaQuaglia M, Moll UM (1998) Expression level, allelic origin, and mutation analysis of the p73 gene in neuroblastoma tumors and cell lines. *Cell Growth Differ* **9**:897-903

Kozasa T, Itoh H, Tsukamoto T, Kaziro Y (1988) Isolation and characterization of the human Gs alpha gene. *Proc Natl Acad Sci U S A* **85**:2081-5

Kroiss MM, Bosserhoff AK, Vogt T, Buettner R, Bogenrieder T, Landthaler M, Stolz W (1998) Loss of expression or mutations in the p73 tumour suppressor gene are not involved in the pathogenesis of malignant melanomas. *Melanoma Res* **8**:504-9

Kubota T, Das S, Christian SL, Baylin SB, Herman JG, Ledbetter DH (1997) Methylation-specific PCR simplifies imprinting analysis. *Nat Genet* **16**:16-7

Kuida K, Zheng TS, Na S, Kuan C, Yang D, Karasuyama H, Rakic P, Flavell RA (1996) Decreased apoptosis in the brain and premature lethality in CPP32-deficient mice. *Nature* **384**:368-72

Lahti JM, Valentine M, Xiang J, Jones B, Amann J, Grenet J, Richmond G, Look AT, Kidd VJ (1994) Alterations in the PITSLRE protein kinase gene complex on chromosome 1p36 in childhood neuroblastoma. *Nat Genet* **7**:370-5

Lastowska MA, Lillington DM, Shelling AN, Cooke I, Gibbons B, Young BD, Ganesan TS (1994) Fluorescence in situ hybridization analysis using cosmid probes to define chromosome 6q abnormalities in ovarian carcinoma cell lines. *Cancer Genet Cytogenet* **77**:99-105

Lau MM, Stewart CE, Liu Z, Bhatt H, Rotwein P, Stewart CL (1994) Loss of the imprinted IGF2/cation-independent mannose 6-phosphate receptor results in fetal overgrowth and perinatal lethality. *Genes Dev* **8**:2953-63

Lawler SD, Pickthall VJ, Fisher RA, Povey S, Evans MW, Szulman AE (1979) Genetic studies of complete and partial hydatidiform moles. *Lancet* **2**:580

Lee MP, Hu RJ, Johnson LA, Feinberg AP (1997) Human KVLQT1 gene shows tissue-specific imprinting and encompasses Beckwith-Wiedemann syndrome chromosomal rearrangements. *Nat Genet* **15**:181-5

Lee MP, Brandenburg S, Landes GM, Adams M, Miller G, Feinberg AP (1999a) Two novel genes in the center of the 11p15 imprinted domain escape genomic imprinting. *Hum Mol Genet* **8**:683-90

Lee MP, DeBaun MR, Mitsuya K, Galonek HL, Brandenburg S, Oshimura M, Feinberg AP (1999b) Loss of imprinting of a paternally expressed transcript, with antisense orientation to KVLQT1, occurs frequently in Beckwith-Wiedemann syndrome and is independent of insulin-like growth factor II imprinting. *Proc Natl Acad Sci U S A* **96**:5203-8

Lee S, Kozlov S, Hernandez L, Chamberlain SJ, Brannan CI, Stewart CL, Wevrick R (2000) Expression and imprinting of MAGEL2 suggest a role in prader-willi syndrome and the homologous murine imprinting phenotype. *Hum Mol Genet* **9**:1813-9

Lee S, Wevrick R (2000) Identification of novel imprinted transcripts in the Prader-Willi syndrome and Angelman syndrome deletion region: further evidence for regional imprinting control. *Am J Hum Genet* **66**:848-58

Leek JP, Carr IM, Bell SM, Markham AF, Lench NJ (1997) Assignment of the DNA fragmentation factor gene (DFFA) to human chromosome bands 1p36.3-p36.2 by in situ hybridization. *Cytogenet Cell Genet* **79**:212-3

Lefebvre L, Viville S, Barton SC, Ishino F, Keverne EB, Surani MA (1998) Abnormal maternal behaviour and growth retardation associated with loss of the imprinted gene Mest. *Nat Genet* **20**:163-9

Lefebvre L, Viville S, Barton SC, Ishino F, Surani MA (1997) Genomic structure and parent-of-origin-specific methylation of Peg1. *Hum Mol Genet* **6**:1907-15

Leighton PA, Saam JR, Ingram RS, Stewart CL, Tilghman SM (1995) An enhancer deletion affects both H19 and Igf2 expression. *Genes Dev* **9**:2079-89

Leighton PA, Saam JR, Ingram RS, Tilghman SM (1996) Genomic imprinting in mice: its function and mechanism. *Biol Reprod* **54**:273-8

Lepretre S, Buchonnet G, Stamatoullas A, Lenain P, Duval C, d'Anjou J, Callat MP, Tilly H, Bastard C (2000) Chromosome abnormalities in peripheral T-cell lymphoma. *Cancer Genet Cytogenet* **117**:71-9

Levine MA, Jap TS, Mauseth RS, Downs RW, Spiegel AM (1986) Activity of the stimulatory guanine nucleotide-binding protein is reduced in erythrocytes from patients with pseudohypoparathyroidism and pseudopseudohypoparathyroidism: biochemical, endocrine, and genetic analysis of Albright's hereditary osteodystrophy in six kindreds. *J Clin Endocrinol Metab* **62**:497-502

Levine MA, Ahn TG, Klupt SF, Kaufman KD, Smallwood PM, Bourne HR, Sullivan KA, Van Dop C (1988) Genetic deficiency of the alpha subunit of the guanine

nucleotide-binding protein Gs as the molecular basis for Albright hereditary osteodystrophy. *Proc Natl Acad Sci U S A* **85**:617-21

Li E, Beard C, Jaenisch R (1993) Role for DNA methylation in genomic imprinting. *Nature* **366**:362-5

Li M, Squire JA, Weksberg R (1997) Molecular genetics of Beckwith-Wiedemann syndrome. *Curr Opin Pediatr* **9**:623-9

Li X, Adam G, Cui H, Sandstedt B, Ohlsson R, Ekstrom TJ (1995) Expression, promoter usage and parental imprinting status of insulin-like growth factor II (IGF2) in human hepatoblastoma: uncoupling of IGF2 and H19 imprinting. *Oncogene* **11**:221-9

Linder D, McCaw BK, Hecht F (1975) Parthenogenic origin of benign ovarian teratomas. *N Engl J Med* **292**:63-6

Liu F, Roth RA (1995) Grb-IR: a SH2-domain-containing protein that binds to the insulin receptor and inhibits its function. *Proc Natl Acad Sci U S A* **92**:10287-91

Liu J, Yu S, Litman D, Chen W, Weinstein LS (2000a) Identification of a methylation imprint mark within the mouse Gnas locus. *Mol Cell Biol* **20**:5808-17

Liu W, Mai M, Yokomizo A, Qian C, Tindall DJ, Smith DI, Thibodeau SN (2000b) Differential expression and allelotyping of the p73 gene in neuroblastoma. *Int J Oncol* **16**:181-5

Liu X, Zou H, Slaughter C, Wang X (1997) DFF, a heterodimeric protein that functions downstream of caspase-3 to trigger DNA fragmentation during apoptosis. *Cell* **89**:175-84

Liu X, Li P, Widlak P, Zou H, Luo X, Garrard WT, Wang X (1998) The 40-kDa subunit of DNA fragmentation factor induces DNA fragmentation and chromatin condensation during apoptosis. *Proc Natl Acad Sci U S A* **95**:8461-6

Luo DF, Buzzetti R, Rotter JI, Maclaren NK, Raffel LJ, Nistico L, Giovannini C, Pozzilli P, Thomson G, She JX (1996) Confirmation of three susceptibility genes to insulin-dependent diabetes mellitus: IDDM4, IDDM5 and IDDM8. *Hum Mol Genet* **5:**693-8

Lyle R, Watanabe D, te Vruchte D, Lerchner W, Smrzka OW, Wutz A, Schageman J, Hahner L, Davies C, Barlow DP (2000) The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. *Nat Genet* **25**:19-21

MacDonald HR, Wevrick R (1997) The necdin gene is deleted in Prader-Willi syndrome and is imprinted in human and mouse. *Hum Mol Genet* **6**:1873-8

Magenis RE, Toth-Fejel S, Allen LJ, Black M, Brown MG, Budden S, Cohen R, Friedman JM, Kalousek D, Zonana J, et al. (1990) Comparison of the 15q deletions in Prader-Willi and Angelman syndromes: specific regions, extent of deletions, parental origin, and clinical consequences. *Am J Med Genet* **35**:333-49

Maher ER, Reik W (2000) Beckwith-Wiedemann syndrome: imprinting in clusters revisited. *J Clin Invest.* **105**:247-52

Mai M, Yokomizo A, Qian C, Yang P, Tindall DJ, Smith DI, Liu W (1998a) Activation of p73 silent allele in lung cancer. *Cancer Res* **58**:2347-9

Mai M, Huang H, Reed C, Qian C, Smith JS, Alderete B, Jenkins R, Smith DI, Liu W (1998b) Genomic organization and mutation analysis of p73 in oligodendrogliomas with chromosome 1 p-arm deletions. *Genomics* **51**:359-63

Mai M, Qian C, Yokomizo A, Tindall DJ, Bostwick D, Polychronakos C, Smith DI, Liu W (1998c) Loss of imprinting and allele switching of p73 in renal cell carcinoma. *Oncogene* **17**:1739-41

Mann MR, Bartolomei MS (2000) Maintaining imprinting. *Nat Genet* **25**:4-5

Mannens M, Slater RM, Heyting C, Bliek J, de Kraker J, Coad N, de Pagter-Holthuizen P, Pearson PL (1988) Molecular nature of genetic changes resulting in loss of heterozygosity of chromosome 11 in Wilms' tumours. *Hum Genet* **81**:41-8

Maris JM, Jensen J, Sulman EP, Beltinger CP, Allen C, Biegel JA, Brodeur GM, White PS (1997) Human Kruppel-related 3 (HKR3): a candidate for the 1p36 neuroblastoma tumour suppressor gene? *Eur J Cancer* **33**:1991-6

Marquis E, Robert JJ, Benezech C, Junien C, Diatloff-Zito C (2000) Variable features of transient neonatal diabetes mellitus with paternal isodisomy of chromosome 6. *Eur J Hum Genet* **8**:137-40

Mars WM, Saunders GF (1990) Chromosomal abnormalities in human breast cancer. *Cancer Metastasis Rev* **9**:35-43

Matsuoka S, Thompson JS, Edwards MC, Bartletta JM, Grundy P, Kalikin LM, Harper JW, Elledge SJ, Feinberg AP (1996) Imprinting of the gene encoding a human cyclin-dependent kinase inhibitor, p57KIP2, on chromosome 11p15. *Proc Natl Acad Sci U S A* **93**:3026-30

Matsuura T, Sutcliffe JS, Fang P, Galjaard RJ, Jiang YH, Benton CS, Rommens JM, Beaudet AL (1997) De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome. *Nat Genet* **15**:74-7

Mazurenko N, Attaleb M, Gritsko T, Semjonova L, Pavlova L, Sakharova O, Kisseljov F (1999) High resolution mapping of chromosome 6 deletions in cervical cancer. *Oncol Rep* **6**:859-63

McCann AH, Miller N, O'Meara A, Pedersen I, Keogh K, Gorey T, Dervan PA (1996) Biallelic expression of the IGF2 gene in human breast disease. *Hum Mol Genet* **5**:1123-7

McDonald LE, Paterson CA, Kay GF (1998) Bisulfite genomic sequencing-derived methylation profile of the xist gene throughout early mouse development. *Genomics* **54**:379-86

McGrath J, Solter D (1984) Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **37**:179-83

Menasce LP, Orphanos V, Santibanez-Koref M, Boyle JM, Harrison CJ (1994) Common region of deletion on the long arm of chromosome 6 in non-Hodgkin's lymphoma and acute lymphoblastic leukaemia. *Genes Chromosomes Cancer* **10**:286-8

Merup M, Moreno TC, Heyman M, Ronnberg K, Grander D, Detlofsson R, Rasool O, Liu Y, Soderhall S, Juliusson G, Gahrton G, Einhorn S (1998) 6q deletions in acute lymphoblastic leukemia and non-Hodgkin's lymphomas. *Blood* **91**:3397-400

Miele ME, McGary CT, Welch DR (1995) SOD2 (MnSOD) does not suppress tumorigenicity or metastasis of human melanoma C8161 cells. *Anticancer Res* **15**:2065-70

Mihara M, Nimura Y, Ichimiya S, Sakiyama S, Kajikawa S, Adachi W, Amano J, Nakagawara A (1999) Absence of mutation of the p73 gene localized at chromosome 1p36.3 in hepatocellular carcinoma. *Br J Cancer* **79**:164-7

Milner RD, Ferguson AW, Naidu SH (1971) Aetiology of transient neonatal diabetes. *Arch Dis Child* **46**:724-6

Mitsuya K, Meguro M, Lee MP, Katoh M, Schulz TC, Kugoh H, Yoshida MA, Niikawa N, Feinberg AP, Oshimura M (1999) LIT1, an imprinted antisense RNA in the human KvLQT1 locus identified by screening for differentially expressed transcripts using monochromosomal hybrids. *Hum Mol Genet* **8**:1209-17

Miyoshi N, Kuroiwa Y, Kohda T, Shitara H, Yonekawa H, Kawabe T, Hasegawa H, Barton SC, Surani MA, Kaneko-Ishino T, Ishino F (1998) Identification of the Meg1/Grb10 imprinted gene on mouse proximal chromosome 11, a candidate for the Silver-Russell syndrome gene. *Proc Natl Acad Sci U S A* **95**:1102-7

Monk D, Wakeling EL, Proud V, Hitchins M, Abu-Amero SN, Stanier P, Preece MA, Moore GE (2000) Duplication of 7p11.2-p13, including GRB10, in Silver-Russell syndrome. *Am J Hum Genet* **66**:36-46

Monk M, Boubelik M, Lehnert S (1987) Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**:371-82

Monni O, Joensuu H, Franssila K, Knuutila S (1996) DNA copy number changes in diffuse large B-cell lymphoma--comparative genomic hybridization study. *Blood* **87**:5269-78

Moore T, Haig D (1991) Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet* **7**:45-9

Moore T (1994) Refusing the ovarian time bomb. *Trends Genet* **10**:347-9

Moore T, Constancia M, Zubair M, Bailleul B, Feil R, Sasaki H, Reik W (1997) Multiple imprinted sense and antisense transcripts, differential methylation and tandem repeats in a putative imprinting control region upstream of mouse Igf2. *Proc Natl Acad Sci U S A* **94**:12509-14

Morrione A, Valentinis B, Resnicoff M, Xu S, Baserga R (1997) The role of mGrb10alpha in insulin-like growth factor I-mediated growth. *J Biol Chem* **272**:26382-7

Moutoussamy S, Renaudie F, Lago F, Kelly PA, Finidori J (1998) Grb10 identified as a potential regulator of growth hormone (GH) signaling by cloning of GH receptor target proteins. *J Biol Chem* **273**:15906-12

Mukae N, Enari M, Sakahira H, Fukuda Y, Inazawa J, Toh H, Nagata S (1998) Molecular cloning and characterization of human caspase-activated DNase. *Proc Natl Acad Sci U S A* **95**:9123-8

Muller S, Zirkel D, Westphal M, Zumkeller W (2000) Genomic imprinting of IGF2 and H19 in human meningiomas. *Eur J Cancer* **36**:651-5

Mutter GL, Stewart CL, Chaponot ML, Pomponio RJ (1993) Oppositely imprinted genes H19 and insulin-like growth factor 2 are coexpressed in human androgenetic trophoblast. *Am J Hum Genet* **53**:1096-102

Nagase T, Ishikawa K, Miyajima N, Tanaka A, Kotani H, Nomura N, Ohara O (1998) Prediction of the coding sequences of unidentified human genes. IX. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro. *DNA Res* **5**:31-9

Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**:386-9

Negrini M, Sabbioni S, Possati L, Rattan S, Corallini A, Barbanti-Brodano G, Croce CM (1994) Suppression of tumorigenicity of breast cancer cells by microcell-

mediated chromosome transfer: studies on chromosomes 6 and 11. *Cancer Res* **54**:1331-6

Ng SW, Yiu GK, Liu Y, Huang LW, Palnati M, Jun SH, Berkowitz RS, Mok SC (2000) Analysis of p73 in human borderline and invasive ovarian tumor. *Oncogene* **19**:1885-90

Nicholls RD (1993) Genomic imprinting and uniparental disomy in Angelman and Prader-Willi syndromes: a review. *Am J Med Genet* **46**:16-25

Nimura Y, Mihara M, Ichimiya S, Sakiyama S, Seki N, Ohira M, Nomura N, Fujimori M, Adachi W, Amano J, He M, Ping YM, Nakagawara A (1998) p73, a gene related to p53, is not mutated in esophageal carcinomas. *Int J Cancer* **78**:437-40

Nishita Y, Yoshida I, Sado T, Takagi N (1996) Genomic imprinting and chromosomal localization of the human MEST gene. *Genomics* **36**:539-42

Nomoto S, Haruki N, Kondo M, Konishi H, Takahashi T (1998) Search for mutations and examination of allelic expression imbalance of the p73 gene at 1p36.33 in human lung cancers. *Cancer Res* **58**:1380-3

Nonomura N, Nishimura K, Miki T, Kanno N, Kojima Y, Yokoyama M, Okuyama A (1997) Loss of imprinting of the insulin-like growth factor II gene in renal cell carcinoma. *Cancer Res* **57**:2575-7

Noviello C, Courjal F, Theillet C (1996) Loss of heterozygosity on the long arm of chromosome 6 in breast cancer: possibly four regions of deletion. *Clin Cancer Res* **2**:1601-6

Oda H, Kume H, Shimizu Y, Inoue T, Ishikawa T (1998) Loss of imprinting of igf2 in renal-cell carcinomas. *Int J Cancer* **75**:343-6

Offit K, Parsa NZ, Gaidano G, Filippa DA, Louie D, Pan D, Jhanwar SC, Dalla-Favera R, Chaganti RS (1993) 6q deletions define distinct clinico-pathologic subsets of non-Hodgkin's lymphoma. *Blood* **82**:2157-62

Ogawa O, Eccles MR, Szeto J, McNoe LA, Yun K, Maw MA, Smith PJ, Reeve AE (1993) Relaxation of insulin-like growth factor II gene imprinting implicated in Wilms' tumour. *Nature* **362**:749-51

Ohama K, Kajii T, Okamoto E, Fukuda Y, Imaizumi K, Tsukahara M, Kobayashi K, Hagiwara K (1981) Dispermic origin of XY hydatidiform moles. *Nature* **292**:551-2

Ohama K, Nomura K, Okamoto E, Fukuda Y, Ihara T, Fujiwara A (1985) Origin of immature teratoma of the ovary. *Am J Obstet Gynecol* **152**:896-900

Ohira M, Kageyama H, Mihara M, Furuta S, Machida T, Shishikura T, Takayasu H, Islam A, Nakamura Y, Takahashi M, Tomioka N, Sakiyama S, Kaneko Y, Toyoda

A, Hattori M, Sakaki Y, Ohki M, Horii A, Soeda E, Inazawa J, Seki N, Kuma H, Nozawa I, Nakagawara A (2000) Identification and characterization of a 500-kb homozygously deleted region at 1p36.2-p36.3 in a neuroblastoma cell line. *Oncogene* **19**:4302-7

Ohta T, Gray TA, Rogan PK, Buiting K, Gabriel JM, Saitoh S, Muralidhar B, Bilienska B, Krajewska-Walasek M, Driscoll DJ, Horsthemke B, Butler MG, Nicholls RD (1999) Imprinting-mutation mechanisms in Prader-Willi syndrome. *Am J Hum Genet* **64**:397-413

Olek A, Oswald J, Walter J (1996) A modified and improved method for bisulphite based cytosine methylation analysis. *Nucleic Acids Res* **24**:5064-6

O'Neill TJ, Rose DW, Pillay TS, Hotta K, Olefsky JM, Gustafson TA (1996) Interaction of a GRB-IR splice variant (a human GRB10 homolog) with the insulin and insulin-like growth factor I receptors. Evidence for a role in mitogenic signaling. *J Biol Chem* **271**:22506-13

Orphanos V, McGown G, Hey Y, Thorncroft M, Santibanez-Koref M, Russell SE, Hickey I, Atkinson RJ, Boyle JM (1995) Allelic imbalance of chromosome 6q in ovarian tumours. *Br J Cancer* **71**:666-9

Ozcelik T, Leff S, Robinson W, Donlon T, Lalande M, Sanjines E, Schinzel A, Francke U (1992) Small nuclear ribonucleoprotein polypeptide N (SNRPN), an expressed gene in the Prader-Willi syndrome critical region. *Nat Genet* **2**:265-9

Pandya A, Braverman N, Pyeritz RE, Ying KL, Kline AD, Falk RE (1995) Interstitial deletion of the long arm of chromosome 6 associated with unusual limb anomalies: report of two new patients and review of the literature. *Am J Med Genet* **59**:38-43

Pedersen IS, Dervan PA, Broderick D, Harrison M, Miller N, Delany E, O'Shea D, Costello P, McGoldrick A, Keating G, Tobin B, Gorey T, McCann A (1999) Frequent loss of imprinting of PEG1/MEST in invasive breast cancer. *Cancer Res* **59**:5449-51

Peng CY, Tsai SL, Yeh CT, Hung SP, Chen MF, Chen TC, Chu CM, Liaw YF (2000) Genetic alterations of p73 are infrequent but may occur in early stage hepatocellular carcinoma. *Anticancer Res* **20**:1487-92

Pfeifer K (2000) Mechanisms of Genomic Imprinting. *Am J Hum Genet* **67**:777-787

Piras G, El Kharroubi A, Kozlov S, Escalante-Alcalde D, Hernandez L, Copeland NG, Gilbert DJ, Jenkins NA, Stewart CL (2000) Zac1 (Lot1), a potential tumor suppressor gene, and the gene for epsilon-sarcoglycan are maternally imprinted genes: identification by a subtractive screen of novel uniparental fibroblast lines. *Mol Cell Biol* **20**:3308-15

Pivnick EK, Qumsiyeh MB, Tharapel AT, Summitt JB, Wilroy RS (1990) Partial duplication of the long arm of chromosome 6: a clinically recognisable syndrome. *J Med Genet* **27**:523-6

Prasad R, Gu Y, Alder H, Nakamura T, Canaani O, Saito H, Huebner K, Gale RP, Nowell PC, Kuriyama K, et al. (1993) Cloning of the ALL-1 fusion partner, the AF-6 gene, involved in acute myeloid leukemias with the t(6;11) chromosome translocation. *Cancer Res* **53**:5624-8

Preece MA, Price SM, Davies V, Clough L, Stanier P, Trembath RC, Moore GE (1997) Maternal uniparental disomy 7 in Silver-Russell syndrome. *J Med Genet* **34**:6-9

Queimado L, Reis A, Fonseca I, Martins C, Lovett M, Soares J, Parreira L (1998) A refined localization of two deleted regions in chromosome 6q associated with salivary gland carcinomas. *Oncogene* **16**:83-8

Rabin M, Birnbaum D, Young D, Birchmeier C, Wigler M, Ruddle FH (1987) Human ros1 and mas1 oncogenes located in regions of chromosome 6 associated with tumor-specific rearrangements. *Oncogene Res* **1**:169-78

Rainier S, Johnson LA, Dobry CJ, Ping AJ, Grundy PE, Feinberg AP (1993) Relaxation of imprinted genes in human cancer. *Nature* **362**:747-9

Rainier S, Dobry CJ, Feinberg AP (1995) Loss of imprinting in hepatoblastoma. *Cancer Res* **55**:1836-8

Razin A, Shemer R (1995) DNA methylation in early development. *Hum Mol Genet* **4**:1751-5

Reeve AE, Sih SA, Raizis AM, Feinberg AP (1989) Loss of allelic heterozygosity at a second locus on chromosome 11 in sporadic Wilms' tumor cells. *Mol Cell Biol* **9**:1799-803

Reik W, Brown KW, Slatter RE, Sartori P, Elliott M, Maher ER (1994) Allelic methylation of H19 and IGF2 in the Beckwith-Wiedemann syndrome. *Hum Mol Genet* **3**:1297-301

Reik W, Constancia M (1997) Genomic imprinting. Making sense or antisense? *Nature* **389**:669-71

Reik W, Maher ER (1997) Imprinting in clusters: lessons from Beckwith-Wiedemann syndrome. *Trends Genet* **13**:330-4

Reik W, Walter J (1998) Imprinting mechanisms in mammals. *Curr Opin Genet Dev* **8**:154-64

Richter J, Beffa L, Wagner U, Schraml P, Gasser TC, Moch H, Mihatsch MJ, Sauter G (1998) Patterns of chromosomal imbalances in advanced urinary bladder cancer detected by comparative genomic hybridization. *Am J Pathol* **153**:1615-21

Richter J, Wagner U, Schraml P, Maurer R, Alund G, Knonagel H, Moch H, Mihatsch MJ, Gasser TC, Sauter G (1999) Chromosomal imbalances are associated with a high risk of progression in early invasive (pT1) urinary bladder cancer. *Cancer Res* **59**:5687-91

Rickert CH, Dockhorn-Dworniczak B, Simon R, Paulus W (1999) Chromosomal imbalances in primary lymphomas of the central nervous system. *Am J Pathol* **155**:1445-51

Riesewijk AM, Schepens MT, Welch TR, van den Berg-Loonen EM, Mariman EM, Ropers HH, Kalscheuer VM (1996) Maternal-specific methylation of the human IGF2R gene is not accompanied by allele-specific transcription. *Genomics* **31**:158-66

Riesewijk AM, Hu L, Schulz U, Tariverdian G, Hoglund P, Kere J, Ropers HH, Kalscheuer VM (1997) Monoallelic expression of human PEG1/MEST is paralleled by parent-specific methylation in fetuses. *Genomics* **42**:236-44

Riesewijk AM, Blagitko N, Schinzel AA, Hu L, Schulz U, Hamel BC, Ropers HH, Kalscheuer VM (1998) Evidence against a major role of PEG1/MEST in Silver-Russell syndrome. *Eur J Hum Genet* **6**:114-20

Rodriguez C, Causse A, Ursule E, Theillet C (2000) At least five regions of imbalance on 6q in breast tumors, combining losses and gains. *Genes Chromosomes Cancer* **27**:76-84

Rougeulle C, Glatt H, Lalande M (1997) The Angelman syndrome candidate gene, UBE3A/E6-AP, is imprinted in brain. *Nat Genet.* **17**(1):14-5

Rougeulle C, Cardoso C, Fontes M, Colleaux L, Lalande M (1998) An imprinted antisense RNA overlaps UBE3A and a second maternally expressed transcript. *Nat Genet* **19**:15-6

Russell A (1954) A syndrome of "Intra-uterine" dwarfism recognisable at birth with cranio-facial dysostosis, disproportionately short arms, and other anomalies (5 examples). *Proc R Soc Med* **47**:1040-1044

Saito S, Saito H, Koi S, Sagae S, Kudo R, Saito J, Noda K, Nakamura Y (1992) Fine-scale deletion mapping of the distal long arm of chromosome 6 in 70 human ovarian cancers. *Cancer Res* **52**:5815-7

Saito S, Sirahama S, Matsushima M, Suzuki M, Sagae S, Kudo R, Saito J, Noda K, Nakamura Y (1996) Definition of a commonly deleted region in ovarian cancers to a 300-kb segment of chromosome 6q27. *Cancer Res* **56**:5586-9

Sakahira H, Enari M, Nagata S (1998) Cleavage of CAD inhibitor in CAD activation and DNA degradation during apoptosis. *Nature* **391**:96-9

Sasaki H, Jones PA, Chaillet JR, Ferguson-Smith AC, Barton SC, Reik W, Surani MA (1992) Parental imprinting: potentially active chromatin of the repressed maternal allele of the mouse insulin-like growth factor II (Igf2) gene. *Genes Dev* **6**:1843-56

Schittek B, Sauer B, Garbe C (1999) Lack of p73 mutations and late occurrence of p73 allelic deletions in melanoma tissues and cell lines. *Int J Cancer* **82**:583-6

Schlegelberger B, Himmler A, Bartles H, Kuse R, Sterry W, Grote W (1994) Recurrent chromosome abnormalities in peripheral T-cell lymphomas. *Cancer Genet Cytogenet* **78**:15-22

Schleiermacher G, Peter M, Michon J, Hugot JP, Vielh P, Zucker JM, Magdelenat H, Thomas G, Delattre O (1994) Two distinct deleted regions on the short arm of chromosome 1 in neuroblastoma. *Genes Chromosomes Cancer* **10**:275-81

Schmidt JV, Levorse JM, Tilghman SM (1999) Enhancer competition between H19 and Igf2 does not mediate their imprinting. *Proc Natl Acad Sci U S A* **96**:9733-8

Schroeder WT, Chao LY, Dao DD, Strong LC, Pathak S, Riccardi V, Lewis WH, Saunders GF (1987) Nonrandom loss of maternal chromosome 11 alleles in Wilms tumors. *Am J Hum Genet* **40**:413-20

Schwab M, Alitalo K, Klempnauer KH, Varmus HE, Bishop JM, Gilbert F, Brodeur G, Goldstein M, Trent J (1983) Amplified DNA with limited homology to myc cellular oncogene is shared by human neuroblastoma cell lines and a neuroblastoma tumour. *Nature* **305**:245-8

Schwab M (1990) Amplification of the MYCN oncogene and deletion of putative tumour suppressor gene in human neuroblastomas. *Brain Pathol* **1**:41-6

Schwartz DI, Lindor NM, Walsh-Vockley C, Roche PC, Mai M, Smith DI, Liu W, Couch FJ (1999) p73 mutations are not detected in sporadic and hereditary breast cancer. *Breast Cancer Res Treat* **58**:25-9

Schwienbacher C, Gramantieri L, Scelfo R, Veronese A, Calin GA, Bolondi L, Croce CM, Barbanti-Brodano G, Negrini M (2000) Gain of imprinting at chromosome 11p15: A pathogenetic mechanism identified in human hepatocarcinomas. *Proc Natl Acad Sci U S A* **97**:5445-9

Scrable H, Cavenee W, Ghavimi F, Lovell M, Morgan K, Sapienza C (1989) A model for embryonal rhabdomyosarcoma tumorigenesis that involves genome imprinting. *Proc Natl Acad Sci U S A* **86**:7480-4

Seeger RC, Brodeur GM, Sather H, Dalton A, Siegel SE, Wong KY, Hammond D (1985) Association of multiple copies of the N-myc oncogene with rapid progression of neuroblastomas. *N Engl J Med* **313**:1111-6

Shemer R, Birger Y, Riggs AD, Razin A (1997) Structure of the imprinted mouse Snrpn gene and establishment of its parental-specific methylation pattern. *Proc Natl Acad Sci U S A* **94**:10267-72

Sheng ZM, Marchetti A, Buttitta F, Champeme MH, Campani D, Bistocchi M, Lidereau R, Callahan R (1996) Multiple regions of chromosome 6q affected by loss of heterozygosity in primary human breast carcinomas. *Br J Cancer* **73**:144-7

Sherratt T, Morelli C, Boyle JM, Harrison CJ (1997) Analysis of chromosome 6 deletions in lymphoid malignancies provides evidence for a region of minimal deletion within a 2-megabase segment of 6q21. *Chromosome Res* **5**:118-24

Shibata H, Ueda T, Kamiya M, Yoshiki A, Kusakabe M, Plass C, Held WA, Sunahara S, Katsuki M, Muramatsu M, Hayashizaki Y (1997) An oocyte-specific methylation imprint center in the mouse U2afbp-rs/U2af1-rs1 gene marks the establishment of allele-specific methylation during preimplantation development. *Genomics* **44**:171-8

Shield JP, Gardner RJ, Wadsworth EJ, Whiteford ML, James RS, Robinson DO, Baum JD, Temple IK (1997) Aetiopathology and genetic basis of neonatal diabetes. *Arch Dis Child Fetal Neonatal Ed* **76**:F39-42

Shishikura T, Ichimiya S, Ozaki T, Nimura Y, Kageyama H, Nakamura Y, Sakiyama S, Miyauchi M, Yamamoto N, Suzuki M, Nakajima N, Nakagawara A (1999) Mutational analysis of the p73 gene in human breast cancers. *Int J Cancer* **84**:321-5

Shridhar V, Staub J, Huntley B, Cliby W, Jenkins R, Pass HI, Hartmann L, Smith DI (1999) A novel region of deletion on chromosome 6q23.3 spanning less than 500 Kb in high grade invasive epithelial ovarian cancer. *Oncogene* **18**:3913-8

Silver HK, Kiyasu W, George J, Dreamer WC (1953) Syndrome of congenital hemihypertrophy, shortness of stature, and elevated urinary gonadotropins. *Pediatrics* **12**:368-375

Smilinich NJ, Day CD, Fitzpatrick GV, Caldwell GM, Lossie AC, Cooper PR, Smallwood AC, Joyce JA, Schofield PN, Reik W, Nicholls RD, Weksberg R, Driscoll DJ, Maher ER, Shows TB, Higgins MJ (1999) A maternally methylated CpG island in KvLQT1 is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. *Proc Natl Acad Sci U S A* **96**:8064-9

Smrzka OW, Fae I, Stoger R, Kurzbauer R, Fischer GF, Henn T, Weith A, Barlow DP (1995) Conservation of a maternal-specific methylation signal at the human IGF2R locus. *Hum Mol Genet* **4**:1945-52

Solter D (1994) Refusing the ovarian time bomb. *Trends Genet* **10**:346; discussion 348-9

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**:503-17

Speel EJ, Richter J, Moch H, Egenter C, Saremaslani P, Rutimann K, Zhao J, Barghorn A, Roth J, Heitz PU, Komminoth P (1999) Genetic differences in endocrine pancreatic tumor subtypes detected by comparative genomic hybridization. *Am J Pathol* **155**:1787-94

Spengler D, Villalba M, Hoffmann A, Pantaloni C, Houssami S, Bockaert J, Journot L (1997) Regulation of apoptosis and cell cycle arrest by Zac1, a novel zinc finger protein expressed in the pituitary gland and the brain. *Embo J* **16**:2814-25

Srikantan V, Sesterhenn IA, Davis L, Hankins GR, Avallone FA, Livezey JR, Connelly R, Mostofi FK, McLeod DG, Moul JW, Chandrasekharappa SC, Srivastava S (1999) Allelic loss on chromosome 6Q in primary prostate cancer. *Int J Cancer* **84**:331-5

Starostik P, Greiner A, Schultz A, Zettl A, Peters K, Rosenwald A, Kolve M, Muller-Hermelink HK (2000) Genetic aberrations common in gastric high-grade large B-cell lymphoma. *Blood* **95**:1180-7

Steenman MJ, Rainier S, Dobry CJ, Grundy P, Horon IL, Feinberg AP (1994) Loss of imprinting of IGF2 is linked to reduced expression and abnormal methylation of H19 in Wilms' tumour. *Nat Genet* **7**:433-9

Strain L, Warner JP, Johnston T, Bonthron DT (1995) A human parthenogenetic chimaera. *Nat Genet* **11**:164-9

Sullivan MJ, Taniguchi T, Jhee A, Kerr N, Reeve AE (1999) Relaxation of IGF2 imprinting in Wilms tumours associated with specific changes in IGF2 methylation. *Oncogene* **18**:7527-34

Sunahara M, Ichimiya S, Nimura Y, Takada N, Sakiyama S, Sato Y, Todo S, Adachi W, Amano J, Nakagawara A (1998) Mutational analysis of the p73 gene localized at chromosome 1p36.3 in colorectal carcinomas. *Int J Oncol* **13**:319-23

Surani MA, Barton SC, Norris ML (1984) Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308**:548-50

Sutcliffe JS, Jiang YH, Galijaard RJ, Matsuura T, Fang P, Kubota T, Christian SL, Bressler J, Cattanach B, Ledbetter DH, Beaudet AL (1997) The E6-Ap ubiquitin-protein ligase (UBE3A) gene is localized within a narrowed Angelman syndrome critical region. *Genome Res* **7**:368-77

Sutcliffe JS, Nakao M, Christian S, Orstavik KH, Tommerup N, Ledbetter DH, Beaudet AL (1994) Deletions of a differentially methylated CpG island at the SNRPN gene define a putative imprinting control region. *Nat Genet* **8**:52-8

Suzuki H, Veda R, Takahashi T (1994) Altered imprinting in lung cancer. *Nat Genet* **6**:332-3

Suzuki M, Saito S, Saga Y, Ohwada M, Sato I (1998) Loss of heterozygosity on chromosome 6q27 and p53 mutations in epithelial ovarian cancer. *Med Oncol* **15**:119-23

Szabo PE, Mann JR (1995) Biallelic expression of imprinted genes in the mouse germ line: implications for erasure, establishment, and mechanisms of genomic imprinting. *Genes Dev* **9**:1857-68

Szabo P, Tang SH, Rentsendorj A, Pfeifer GP, Mann JR (2000) Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. *Curr Biol* **10**:607-10

Takahashi H, Ichimiya S, Nimura Y, Watanabe M, Furusato M, Wakui S, Yatani R, Aizawa S, Nakagawara A (1998) Mutation, allelotyping, and transcription analyses of the p73 gene in prostatic carcinoma. *Cancer Res* **58**:2076-7

Takayama H, Suzuki T, Mugishima H, Fujisawa T, Ookuni M, Schwab M, Gehring M, Nakamura Y, Sugimura T, Terada M, et al. (1992) Deletion mapping of chromosomes 14q and 1p in human neuroblastoma. *Oncogene* **7**:1185-9

Takeda O, Homma C, Maseki N, Sakurai M, Kanda N, Schwab M, Nakamura Y, Kaneko Y (1994) There may be two tumor suppressor genes on chromosome arm 1p closely associated with biologically distinct subtypes of neuroblastoma. *Genes Chromosomes Cancer* **10**:30-9

Tannapfel A, Wasner M, Krause K, Geissler F, Katalinic A, Hauss J, Mossner J, Engeland K, Wittekind C (1999) Expression of p73 and its relation to histopathology and prognosis in hepatocellular carcinoma. *J Natl Cancer Inst* **91**:1154-8

Temple IK, Gardner RJ, Robinson DO, Kibirige MS, Ferguson AW, Baum JD, Barber JC, James RS, Shield JP (1996) Further evidence for an imprinted gene for neonatal diabetes localised to chromosome 6q22-q23. *Hum Mol Genet* **5**:1117-21

Temple IK, James RS, Crolla JA, Sitch FL, Jacobs PA, Howell WM, Betts P, Baum JD, Shield JP (1995) An imprinted gene(s) for diabetes? *Nat Genet* **9**:110-2

Theile M, Seitz S, Arnold W, Jandrig B, Frege R, Schlag PM, Haensch W, Guski H, Winzer KJ, Barrett JC, Scherneck S (1996) A defined chromosome 6q fragment (at D6S310) harbors a putative tumor suppressor gene for breast cancer. *Oncogene* **13**:677-85

Thomas JH (1995) Genomic imprinting proposed as a surveillance mechanism for chromosome loss. *Proc Natl Acad Sci U S A* **92**:480-2

Thorvaldsen JL, Duran KL, Bartolomei MS (1998) Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. *Genes Dev* **12**:3693-702

Tibiletti MG, Bernasconi B, Furlan D, Riva C, Trubia M, Buraggi G, Franchi M, Bolis P, Mariani A, Frigerio L, Capella C, Taramelli R (1996) Early involvement of 6q in surface epithelial ovarian tumors. *Cancer Res* **56**:4493-8

Tibiletti MG, Trubia M, Ponti E, Sessa L, Acquati F, Furlan D, Bernasconi B, Fichera M, Mihalich A, Ziegler A, Volz A, Facco C, Riva C, Cremonesi L, Ferrari M, Taramelli R (1998) Physical map of the D6S149-D6S193 region on chromosome 6Q27 and its involvement in benign surface epithelial ovarian tumours. *Oncogene* **16**:1639-42

Tokuchi Y, Hashimoto T, Kobayashi Y, Hayashi M, Nishida K, Hayashi S, Imai K, Nakachi K, Ishikawa Y, Nakagawa K, Kawakami Y, Tsuchiya E (1999) The expression of p73 is increased in lung cancer, independent of p53 gene alteration. *Br J Cancer* **80**:1623-9

Trask BJ, van den Engh G, Christensen M, Massa HF, Gray JW, Van Dilla M (1991) Characterization of somatic cell hybrids by bivariate flow karyotyping and fluorescence in situ hybridization. *Somat Cell Mol Genet* **17**:117-36

Tremblay KD, Saam JR, Ingram RS, Tilghman SM, Bartolomei MS (1995) A paternal-specific methylation imprint marks the alleles of the mouse H19 gene. *Nat Genet* **9**:407-13

Tremblay KD, Duran KL, Bartolomei MS (1997) A 5' 2-kilobase-pair region of the imprinted mouse H19 gene exhibits exclusive paternal methylation throughout development. *Mol Cell Biol* **17**:4322-9

Trent JM, Thompson FH, Meyskens FL, Jr. (1989) Identification of a recurring translocation site involving chromosome 6 in human malignant melanoma. *Cancer Res* **49**:420-3

Tsao H, Zhang X, Majewski P, Haluska FG (1999) Mutational and expression analysis of the p73 gene in melanoma cell lines. *Cancer Res* **59**:172-4

Tsui SK, Chan JY, Waye MM, Fung KP, Lee CY (1996) Identification of a cDNA encoding 6-phosphogluconate dehydrogenase from a human heart cDNA library. *Biochem Genet* **34**:367-73

Tsujimoto T, Mochizuchi S, Iwadate Y, Namba H, Nagai M, Kawamoto T, Sunahara M, Yamaura A, Nakagawara A, Sakiyama S, Tagawa M (2000) The p73 gene is not

mutated in oligodendrogliomas which frequently have a deleted region at chromosome 1p36.3. *Anticancer Res* **20**:2495-7

Ueda T, Abe K, Miura A, Yuzuriha M, Zubair M, Noguchi M, Niwa K, Kawase Y, Kono T, Matsuda Y, Fujimoto H, Shibata H, Hayashizaki Y, Sasaki H (2000) The paternal methylation imprint of the mouse H19 locus is acquired in the gonocyte stage during foetal testis development. *Genes Cells* **5**:649-59

Utada Y, Haga S, Kajiwara T, Kasumi F, Sakamoto G, Nakamura Y, Emi M (2000) Mapping of target regions of allelic loss in primary breast cancers to 1-cM intervals on genomic contigs at 6q21 and 6q25.3. Jpn J *Cancer Res* **91**:293-300

Van Gele M, Van Roy N, Jauch A, Laureys G, Benoit Y, Schelfhout V, De Potter CR, Brock P, Uyttebroeck A, Sciot R, Schuuring E, Versteeg R, Speleman F (1997) Sensitive and reliable detection of genomic imbalances in human neuroblastomas using comparative genomic hybridisation analysis. *Eur J Cancer* **33**:1979-82

Van Gele M, Speleman F, Vandesompele J, Van Roy N, Leonard JH (1998a) Characteristic pattern of chromosomal gains and losses in Merkel cell carcinoma detected by comparative genomic hybridization. *Cancer Res* **58**:1503-8

Van Gele M, Van Roy N, Ronan SG, Messiaen L, Vandesompele J, Geerts ML, Naeyaert JM, Blennow E, Bar-Am I, Das Gupta TK, van der Drift P, Versteeg R, Leonard JH, Speleman F (1998b) Molecular analysis of 1p36 breakpoints in two Merkel cell carcinomas. *Genes Chromosomes Cancer* **23**:67-71

Van Gele M, Kaghad M, Leonard JH, Van Roy N, Naeyaert JM, Geerts ML, Van Belle S, Cocquyt V, Bridge J, Sciot R, De Wolf-Peeters C, De Paepe A, Caput D, Speleman F (2000) Mutation analysis of P73 and TP53 in Merkel cell carcinoma. *Br J Cancer* **82**:823-6

Van Gurp RJ, Oosterhuis JW, Kalscheuer V, Mariman EC, Looijenga LH (1994) Biallelic expression of the H19 and IGF2 genes in human testicular germ cell tumors. *J Natl Cancer Inst* **86**:1070-5

Varmuza S (1993) Gametic imprinting as a speciation mechanism in mammals. *J Theor Biol* **164**:1-13

Varmuza S, Mann M (1994) Genomic imprinting--defusing the ovarian time bomb. *Trends Genet* **10**:118-23

Varrault A, Ciani E, Apiou F, Bilanges B, Hoffmann A, Pantaloni C, Bockaert J, Spengler D, Journot L (1998) hZAC encodes a zinc finger protein with antiproliferative properties and maps to a chromosomal region frequently lost in cancer. *Proc Natl Acad Sci U S A* **95**:8835-40

Vikhanskaya F, D'Incalci M, Broggini M (2000) p73 competes with p53 and attenuates its response in a human ovarian cancer cell line. *Nucleic Acids Res* **28**:513-9

Villar AJ, Pedersen RA (1994) Parental imprinting of the Mas protooncogene in mouse. *Nat Genet* **8**:373-9

Visakorpi T, Kallioniemi AH, Syvanen AC, Hyytinen ER, Karhu R, Tammela T, Isola JJ, Kallioniemi OP (1995) Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization. *Cancer Res* **55**:342-7

Von Muhlendahl KE, Herkenhoff H (1995) Long-term course of neonatal diabetes. *N Engl J Med* **333**:704-8

Vu TH, Hoffman AR (1994) Promoter-specific imprinting of the human insulin-like growth factor-II gene. *Nature* **371**:714-7

Vu TH, Hoffman AR (1997) Imprinting of the Angelman syndrome gene, UBE3A, is restricted to brain. *Nat Genet* **17**:12-3

Wakeling EL, Hitchins MP, Abu-Amero SN, Stanier P, Moore GE, Preece MA (2000) Biallelic expression of IGFBP1 and IGFBP3, two candidate genes for the Silver-Russell syndrome. *J Med Genet* **37**:65-7

Wan M, Sun T, Vyas R, Zheng J, Granada E, Dubeau L (1999) Suppression of tumorigenicity in human ovarian cancer cell lines is controlled by a 2 cM fragment in chromosomal region 6q24-q25. *Oncogene* **18**:1545-51

Warnecke PM, Mann JR, Frommer M, Clark SJ (1998) Bisulfite sequencing in preimplantation embryos: DNA methylation profile of the upstream region of the mouse imprinted H19 gene. *Genomics* **51**:182-90

Watanabe T, Yoshimura A, Mishima Y, Endo Y, Shiroishi T, Koide T, Sasaki H, Asakura H, Kominami R (2000) Differential chromatin packaging of genomic imprinted regions between expressed and non-expressed alleles. *Hum Mol Genet* **9**:3029-35

Weber T, Weber RG, Kaulich K, Actor B, Meyer-Puttlitz B, Lampel S, Buschges R, Weigel R, Deckert-Schluter M, Schmiedek P, Reifenberger G, Lichter P (2000) Characteristic chromosomal imbalances in primary central nervous system lymphomas of the diffuse large B-cell type. *Brain Pathol* **10**:73-84

Weith A, Martinsson T, Cziepluch C, Bruderlein S, Amler LC, Berthold F, Schwab M (1989) Neuroblastoma consensus deletion maps to 1p36.1-2. *Genes Chromosomes Cancer* **1**:159-66

Wevrick R, Kerns JA, Francke U (1994) Identification of a novel paternally expressed gene in the Prader-Willi syndrome region. *Hum Mol Genet* **3**:1877-82

White PS, Kaufman BA, Marshall HN, Brodeur GM (1993) Use of the single-strand conformation polymorphism technique to detect loss of heterozygosity in neuroblastoma. *Genes Chromosomes Cancer* **7**:102-8

White PS, Maris JM, Beltinger C, Sulman E, Marshall HN, Fujimori M, Kaufman BA, Biegel JA, Allen C, Hilliard C, et al. (1995) A region of consistent deletion in neuroblastoma maps within human chromosome 1p36.2-36.3. *Proc Natl Acad Sci USA* **92**:5520-4

Will GK, Soukupova M, Hong X, Erdmann KS, Kiel JA, Dodt G, Kunau WH, Erdmann R (1999) Identification and characterization of the human orthologue of yeast Pex14p. *Mol Cell Biol* **19**:2265-77

Williams CA, Zori RT, Stone JW, Gray BA, Cantu ES, Ostrer H (1990) Maternal origin of 15q11-13 deletions in Angelman syndrome suggests a role for genomic imprinting. *Am J Med Genet* **35**:350-3

Williams JC, Brown KW, Mott MG, Maitland NJ (1989) Maternal allele loss in Wilms' tumour. *Lancet* **1**:283-4

Wilson S (1991) Transient neonatal diabetes. *Nurs Times* **87**:44-5

Wolffe AP (2000) Imprinting insulation. *Curr Biol* **10**:R463-5

Wroe SF, Kelsey G, Skinner JA, Bodle D, Ball ST, Beechey CV, Peters J, Williamson CM (2000) An imprinted transcript, antisense to Nesp, adds complexity to the cluster of imprinted genes at the mouse Gnas locus. *Proc Natl Acad Sci U S A* **97**:3342-6

Wu SQ, Storer BE, Bookland EA, Klingelhutz AJ, Gilchrist KW, Meisner LF, Oyasu R, Reznikoff CA (1991) Nonrandom chromosome losses in stepwise neoplastic transformation in vitro of human uroepithelial cells. *Cancer Res* **51**:3323-6

Wu HK, Squire JA, Catzavelos CG, Weksberg R (1997a) Relaxation of imprinting of human insulin-like growth factor II gene, IGF2, in sporadic breast carcinomas. *Biochem Biophys Res Commun* **235**:123-9

Wu MS, Wang HP, Lin CC, Sheu JC, Shun CT, Lee WJ, Lin JT (1997b) Loss of imprinting and overexpression of IGF2 gene in gastric adenocarcinoma. *Cancer Lett* **120**:9-14

Wu HK, Weksberg R, Minden MD, Squire JA (1997c) Loss of imprinting of human insulin-like growth factor II gene, IGF2, in acute myeloid leukemia. *Biochem Biophys Res Commun* **231**:466-72

Wutz A, Smrzka OW, Schweifer N, Schellander K, Wagner EF, Barlow DP (1997) Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature* **389**:745-9

Xu Y, Goodyer CG, Deal C, Polychronakos C (1993) Functional polymorphism in the parental imprinting of the human IGF2R gene. *Biochem Biophys Res Commun* **197**:747-54

Xu YQ, Grundy P, Polychronakos C (1997) Aberrant imprinting of the insulin-like growth factor II receptor gene in Wilms' tumor. *Oncogene* **14**:1041-6

Yamada T, De Souza AT, Finkelstein S, Jirtle RL (1997) Loss of the gene encoding mannose 6-phosphate/insulin-like growth factor II receptor is an early event in liver carcinogenesis. *Proc Natl Acad Sci U S A* **94**:10351-5

Yamasaki K, Hayashida S, Miura K, Masuzaki H, Ishimaru T, Niikawa N, Kishino T (2000) The novel gene, gamma2-COP (COPG2), in the 7q32 imprinted domain escapes genomic imprinting. *Genomics* **68**:330-5

Yang A, Walker N, Bronson R, Kaghad M, Oosterwegel M, Bonnin J, Vagner C, Bonnet H, Dikkes P, Sharpe A, McKeon F, Caput D (2000) p73-deficient mice have neurological, pheromonal and inflammatory defects but lack spontaneous tumours. *Nature* **404**:99-103

Yang T, Adamson TE, Resnick JL, Leff S, Wevrick R, Francke U, Jenkins NA, Copeland NG, Brannan CI (1998) A mouse model for Prader-Willi syndrome imprinting-centre mutations. *Nat Genet* **19**:25-31

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**:103-19

Yokomizo A, Mai M, Bostwick DG, Tindall DJ, Qian J, Cheng L, Jenkins RB, Smith DI, Liu W (1999a) Mutation and expression analysis of the p73 gene in prostate cancer. *Prostate* **39**:94-100

Yokomizo A, Mai M, Tindall DJ, Cheng L, Bostwick DG, Naito S, Smith DI, Liu W (1999b) Overexpression of the wild type p73 gene in human bladder cancer. *Oncogene* **18**:1629-33

Yoshihashi H, Maeyama K, Kosaki R, Ogata T, Tsukahara M, Goto Y, Hata J, Matsuo N, Smith RJ, Kosaki K (2000) Imprinting of human GRB10 and its mutations in two patients with russell-silver syndrome. *Am J Hum Genet* **67**:476-82

Yoshikawa H, Nagashima M, Khan MA, McMenamin MG, Hagiwara K, Harris CC (1999) Mutational analysis of p73 and p53 in human cancer cell lines. *Oncogene* **18**:3415-21

Young D, Waitches G, Birchmeier C, Fasano O, Wigler M (1986) Isolation and characterization of a new cellular oncogene encoding a protein with multiple potential transmembrane domains. *Cell* **45**:711-9

Yu Y, Xu F, Peng H, Fang X, Zhao S, Li Y, Cuevas B, Kuo WL, Gray JW, Siciliano M, Mills GB, Bast RC, Jr. (1999) NOEY2 (ARHI), an imprinted putative tumor suppressor gene in ovarian and breast carcinomas. *Proc Natl Acad Sci U S A* **96**:214-9

Zaika AI, Kovalev S, Marchenko ND, Moll UM (1999) Overexpression of the wild type p73 gene in breast cancer tissues and cell lines. *Cancer Res* **59**:3257-63

Zhan S, Shapiro DN, Helman LJ (1995) Loss of imprinting of IGF2 in Ewing's sarcoma. *Oncogene* **11**:2503-7

Zhang J, Liu X, Scherer DC, van Kaer L, Wang X, Xu M (1998a) Resistance to DNA fragmentation and chromatin condensation in mice lacking the DNA fragmentation factor 45. *Proc Natl Acad Sci U S A* **95**:12480-5

Zhang Y, Weber-Matthiesen K, Siebert R, Matthiesen P, Schlegelberger B (1997) Frequent deletions of 6q23-24 in B-cell non-Hodgkin's lymphomas detected by fluorescence in situ hybridization. *Genes Chromosomes Cancer* **18**:310-3

Zhang Y, Matthiesen P, Siebert R, Harder S, Theile M, Scherneck S, Schlegelberger B (1998b) Detection of 6q deletions in breast carcinoma cell lines by fluorescence in situ hybridization. *Hum Genet* **103**:727-9

Zhang Y, Matthiesen P, Harder S, Siebert R, Castoldi G, Calasanz MJ, Wong KF, Rosenwald A, Ott G, Atkin NB, Schlegelberger B (2000) A 3-cM commonly deleted region in 6q21 in leukemias and lymphomas delineated by fluorescence in situ hybridization. *Genes Chromosomes Cancer* **27**:52-8

APPENDIX

Hum Genet (2000) 106:406–413
Digital Object Identifier (DOI) 10.1007/s004390000257

ORIGINAL INVESTIGATION

**Hannah Judson · Nadine van Roy · Lisa Strain
Jo Vandesompele · Mireille Van Gele · Frank Speleman
David T. Bonthron**

Structure and mutation analysis of the gene encoding DNA fragmentation factor 40 (caspase-activated nuclease), a candidate neuroblastoma tumour suppressor gene

Received: 18 November 1999 / Accepted: 23 January 2000 / Published online: 8 March 2000

Abstract We have characterised the *DFFB* gene, encoding the active subunit of the apoptotic nuclease DNA fragmentation factor (DFF40). *DFFB* maps to 1p36, near the imprinted putative tumour suppressor gene *TP73*. The *DFFA* gene (encoding the inhibitory DFF45 subunit) also maps to 1p36.2–36.3, and we show by FISH that *DFFB* lies distal to *DFFA*. We have also mapped a processed *DFFB* pseudogene to chromosome 9. *DFFB* itself has seven coding exons spanning 10 kb. Exhaustive mutation screening of 41 neuroblastomas and other tumours in which a 1p36 tumour suppressor gene is implicated showed no tumour-specific mutations. A coding region polymorphism was used to demonstrate uniformly biallelic expression in human fetal *DFFB* transcripts. Since the putative neuroblastoma tumour suppressor gene in distal 1p36 is predicted to be maternally expressed, the lack of imprinting and absence of somatic mutations in *DFFB* indicate that it is probably not the neuroblastoma tumour suppressor gene.

Introduction

Neuroblastoma is one of the most common childhood tumours, accounting for 8% of all childhood cancers. There is a wide range of outcomes, from cases in which the tumour regresses, with excellent survival prospects, to those in which the disease is much more aggressive and the prognosis poor (Brodeur 1995). A number of chromosomal abnormalities specifically associated with neuroblastoma have been identified by classical cytogenetics, loss of heterozygosity (LOH) analyses, and more recently by comparative genomic hybridisation (Brodeur and Fong 1989; Takita et al. 1997; Vandesompele et al. 1998). One abnormality correlating strongly with a poor outcome is deletion of chromosome 1p (Caron et al. 1996; Maris et al. 1995). Analysis of such deletions suggests that region 1p36 contains more than one tumour suppressor gene (Cheng et al. 1995). From analysis of tumours with *MYCN* amplification, in which 1p deletions are usually large, a tumour suppressor locus is thought to be located at 1p36.1, just distal to *D1S7*. In these tumours, LOH is not selective for one or the other parental allele (Caron et al. 1995). A second commonly deleted region is located at 1p36.2-p36.3. In contrast to the putative 1p36.1 tumour suppressor gene, LOH at this more distal locus usually results from loss of the maternal allele, suggesting the imprinted nature of the tumour suppressor gene (Caron et al. 1993; Caron et al. 1995). Either deletion or mutation of the maternal allele of this gene might be important in the pathogenesis of neuroblastoma. This gene could also be implicated in many other tumours, as 1p deletion has recently been documented as a recurrent aberration in an increasing number of neoplasms. These include common cancers such as melanoma, breast cancer and colon cancer (Schwab et al. 1996) and rare ones such as Merkel cell carcinoma (Van Gele et al. 1998a; Van Gele et al. 1998b) and malignant schwannoma (Van Roy et al., unpublished).

Recently, the *TP73* gene at 1p36 has been proposed as a candidate neuroblastoma tumour suppressor gene (Kaghad et al. 1997). Its protein product appears to have cell cycle control functions analogous to those of the structurally similar p53 protein (Jost et al. 1997). In addition, as predicted for the 1p36.2-p36.3 neuroblastoma tumour suppressor, *TP73* is monoallelically expressed, at least in some tissues (Kaghad et al. 1997; Kovalev et al. 1998; Mai et al. 1998; Nomoto et al. 1998). However, because of the failure to identify somatic mutations in *TP73*, this gene's importance in the pathogenesis of neuroblas-

H. Judson · D.T. Bonthron (✉)
Molecular Medicine Unit, University of Leeds,
St. James's University Hospital, Leeds, LS9 7TF, U.K.
e-mail: meddtb@gps.leeds.ac.uk,
Tel.: +44-1132065681, Fax: +44-1132444475

N. van Roy · J. Vandesompele · M. Van Gele · F. Speleman
Department of Medical Genetics, University Hospital,
1K5, De Pintelaan 185, 9000 Gent, Belgium

L. Strain
Human Genetics Unit, University of Edinburgh,
Western General Hospital, Edinburgh, EH4 2XU, U.K.

toma remains unproven (Kovalev et al. 1998). The identification of other candidate genes in this region is therefore an important goal.

Since imprinted genes are often clustered within chromosomal regions containing several imprinted and non-imprinted genes (Reik and Maher 1997), we have searched for novel imprinted genes in the region surrounding *TP73*. We have thereby identified, at a distance of 100–200 kb from *TP73*, the gene encoding the apoptotic nuclease variously known as caspase-activated nuclease (CPAN) (Halenbeck et al. 1998), caspase-activated DNase (CAD) (Enari et al. 1998; Mukae et al. 1998) or DNA fragmentation factor (DFF40) (Liu et al. 1997). This gene (here referred to as *DFFB*) encodes a 40 kDa nuclease that is activated by caspase 3, and is capable of degrading DNA during apoptosis. DFF40 activity is regulated by its inhibitor, ICAD/DFF45 (Liu et al. 1997; Sakahira et al. 1998). DFF45 exists in a heterodimeric complex with DFF40 and is proteolytically degraded by activated caspase 3.

The chromosomal location and putative biological role of *DFFB* make it an excellent candidate tumour suppressor gene for neuroblastoma and other tumours frequently exhibiting 1p deletions. Loss of *DFFB* activity by deletion or point mutation could cause failure of apoptotic DNA degradation and hence predispose to tumour development. Support for this idea comes from the fact that ablation of caspase 3 function results in an apoptotic defect, which interestingly is confined to the CNS, and results in accumulation of abnormal neuronal precursors (Kuida et al. 1996).

We have therefore determined the genomic organisation of the CPAN/CAD/DFF40 gene (*DFFB*), and addressed the two main questions of whether this gene is mutated in tumours with frequent occurrence of 1p deletions, and whether the gene is imprinted.

Materials and methods

Contig assembly

PAC clones were identified by hybridisation to grids of the RPCI1 library (Ioannou et al. 1994), provided by the UK Human Genome Mapping Project (HGMP) Resource Centre. The probes and conditions were as follows: PAC 79-L8 was identified using a ^{32}P-labelled 482-bp fragment containing exon 2 of *TP73*, generated by PCR using primers as in Kaghad et al. (1997). PCR conditions were 94°C/1 min, 59°C/1 min, 72°C/2 min, for 40 cycles, in 10 mM Tris·HCl/50 mM KCl/12.5 mM KOH/1.5 mM $MgCl_2$/0.1% Triton X-100, in a volume of 50 µl. Overlapping PACs 71-C21 and 286-D6 were identified by hybridisation to oligonucleotide dACAGAAGGCTCGCACTATCGT, derived from the T7 end of PAC 79-L8. Cosmid 176g8 from the LL01NC01 library (Gingrich et al. 1996; Trask et al. 1991) was identified by hybridisation to a 311-bp PCR product from the 5' end of *DFFB*, amplified from genomic DNA. The primers for this PCR were dCCCTTTGACATGGACAGCTGC and dTGCTTCCGCTTCAACCTTGT. Conditions were 94°C/1 min, 57°C/1 min, 72°C/2 min for 40 cycles, in PCR buffer as described above. PAC and cosmid DNAs were extracted using the Qiagen maxiprep kit.

Gene structure

Positions of splice junctions were established by sequencing using the Thermosequenase ^{33}P-terminator cycle sequencing kit (Amersham). The *DFFB* pseudogene was detected through sequencing of PCR products generated from genomic DNA. A pseudogene-specific PCR, using primers dCCCTTTGACATGGACAGCTGC and dATTGCTTCCACCAGTGTAGG, was then used to map the pseudogene on a human monochromosomal somatic cell hybrid panel (Kelsell et al. 1995), obtained from the HGMP Resource Centre. Cycling conditions were 94°C/1 min, 57°C/1 min, 72°C/2 min, for 35 cycles, in the PCR buffer as described above.

FISH on prometaphase and interphase nuclei and fibre-FISH

Probe labelling, in situ hybridisation and immunocytochemical detection were performed as described (Van Roy et al. 1994). Chromosome slides were prepared from synchronised PHA-stimulated peripheral blood lymphocytes (PBL). Fibre-FISH slides were prepared from freshly fixed suspensions of PHA-stimulated PBL as described (Speleman et al. 1997). Images were captured with a high-sensitivity integrated monochrome CCD camera (Sony IMAC-CCD S30) and dedicated software (ISIS, MetaSystems, Germany).

DNA samples and cell lines

In total, 41 DNA samples were screened for *DFFB* mutations. Fourteen neuroblastoma cell lines and five Merkel cell carcinoma cell lines were analysed (for details of their 1p and *MYCN* status, see Table 1). Three stage 1, three stage 2, four stage 4 and one stage 4S neuroblastoma tumours were included (cases 1, 2, 3, 8, 9, 10, 11, 20, 22, 25 and 29 in the study by Vandesompele et al. (1998), where detailed genetic analysis of these tumours is de-

Table 1 Neuroblastoma and Merkel cell carcinoma cell lines. Chromosome 1p status was evaluated by loss of heterozygosity and also by FISH using probes for *D1Z2* (1p36.33) and *D1Z1* (1q12); decreased copy number of *D1Z2* versus *D1Z1* indicates the presence of a 1p deletion. Cell line UISO has an insertion at 1p36.2. (*LOH* loss of heterozygosity, *MNA* amplification of *MYCN*, *ND* not determined)

Cell lines	LOH 1p	*D1Z2* copy	*D1Z1* copy	MNA
Neuroblastoma cell lines				
IMR32	Yes	1	3	Yes
N206	Yes	2	3	Yes
UHG-NP	Yes	2	4	Yes
TR14	Yes	2	3	Yes
SMS-KCNR	Yes	1	2	Yes
STA-NB-3	ND	2	4	Yes
STA-NB-8	ND	1	2	Yes
STA-NB-9	ND	1	2	Yes
STA-NB-10	ND	1	2	Yes
STA-NB-12	ND	1	2	Yes
NGP	t(1;15)	2	2	Yes
SK-N-AS	Yes	1	3	No
GI-ME-N	Yes	2	4	No
SK-N-FI	No	2	2	No
Merkel cell lines				
MCC 13	No	2	2	No
MCC 14/2	t(X;1)	4	4	No
MCC 26	Yes	3	4	No
UISO	No	2	2	No
MKL-1	ND	2	2	No

scribed). Nine Merkel cell carcinoma tumours were investigated, of which two have a known 1p loss. Finally, two malignant schwannomas with 1p deletions (Van Roy et al., in preparation) were included.

Mutation screening

PCR cycling conditions for amplification of each exon were 94°C/1 min, *X*°C/1 min, 72°C/2 min, for 35 cycles where *X*°C is the primer-specific annealing temperature. Forward and reverse PCR primer sequences and *X*°C were as follows: exon 1: dCCAGCTTGCAGAGCTCAC, dGCTGAGGCGAACGAAAACTA, 57°C; exon 2: dTGTAAAACGACGGCCAGTCAGCCTGAGCCTGCTTCTTTA, dCAGGAAACAGCTATGACCTGAGACCCGAGAGTTCACAG, 65°C; exon 3: dTGTAAAACGACGGCCAGTCTCAAGTCTGAGTCCTGGTGATT, dCAGGAAACAGCTATGACCATGAGCACATTTTCTTCCAAGTC, 60°C; exon 4: dCCACTTGGAAGTCTGAGGCAGG, dCTCCCTCACCAGAGCAGCCCCA, 65°C; exon 5: dTGTAAAACGACGGCCAGTTGGAGTGAGATGGATCGAGA, dCAGGAAACAGCTATGACCAGGGCAAGGGCTGAAGGT, 58°C; exon 6: dGGAGGTTGTAGTAAGCCGAGA, dGAAACCCAAGGAGGCAACAC, 58°C; exon 7: dTGTAAAACGACGGCCAGTACTGTGACTGCAATACACTGC, dCAGGAAACAGCTATGACCTTAAAATGATGCCCACGTCA, 60°C.

Mutation screening was performed by direct sequencing using the Thermosequenase ^{33}P-terminator cycle sequencing kit (Amersham). The following nested sequencing primers (sense, antisense) were used: exon 1: dATCTGAGCAGCTGGGCAG, dCCTATTCTCCCCACACGCCT; exon 2: dAGCACAGCTCATTCCGGTCGT, dTGATGGGCACCTGGAGCTAAG; exon 3: dAGGATGTGTCTTCAGCTGGACCG, dATGAGCACATTTTCTTCCAAGTC; exon 4: dAGGACACAGACCCAGACCC, dACAGAGCCTGGCTTCAAAAA; exon 5: dGCGCTGTGCACCAGGCTCACCCA, dGCTGTGTTCCACAGGCCAGC; exon 6: dACTCCAGCCTGGCGACAGAGCGA, dACAGGGCCCTGCAGGCACTCGT; exon 7: dTGCCTGTGGCACTGTCACCACAG, dATGATGCCCACGTCACGCCTCAA.

RT-PCR analysis of *DFFB* imprinting

RT-PCR analysis of *DFFB* imprinting was performed using modifications of previously described methods and materials (Hayward et al. 1998). Initially, fetal genomic DNA-derived exon 7 PCR products were digested with *Ava*I to type the codon 318 *DFFB* polymorphism. 10 µl of PCR product were digested overnight at 37°C, with 30 U of enzyme in a volume of 20 µl, and analysed on 2% agarose gels.

RT-PCR was next performed on RNA samples from heterozygous fetuses with homozygous mothers. RNA was combined with 500 nM of primer dGAGCTCGAGTCGACATCGA$(T)_{17}$, in water, incubated at 85°C for 3 min, and cooled on ice for 2 min. Reverse transcription components were added to give 50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 1 mM each dNTP, 100 U Superscript II (Gibco/BRL), and 39 U RNAguard (Pharmacia), in a total volume of 30 µl. The mix was incubated at 42°C/60 min, 50°C/15 min, 95°C/15 min. For PCR, 2 µl of this RT mix was used in a 50 µl reaction, containing 10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$, 10 pmol each primer (dATCCTCTTCAGCACCTGGAA and dGTTTCCGCACAGGCTGCT), 200 µM each dNTP, 10% DMSO and 1 U Amplitaq Gold (Perkin Elmer). Cycling conditions were 94°C/45 s, 54°C/45 s, 72°C/120 s, for 40 cycles. Digestion of 25 µl of product was performed in a volume of 40 µl, with 40 U *Ava*I, at 37°C for 3 h, and products were analysed on 8% polyacrylamide gels.

Results

Contig

Initially, a PAC clone (79-L8) was isolated, that contains the *TP73* gene. Using an STS at one end of its insert (T7 end), two overlapping PACs (71-C21 and 286-D6) were then identified. The sequence of the non-overlapping end of 71-C21 proved to be identical to part of a cDNA encoding CPAN/CAD/DFF40 (Halenbeck et al. 1998). In fact, this PAC clone (71-C21) has one end within exon 6 of *DFFB*, the gene encoding CPAN/CAD/DFF40 (see below). Finally, therefore, an STS from this end of PAC 71-C21 was used to isolate a cosmid (176g8) containing the complete *DFFB* gene.

FISH

The *DFFA* gene encodes DFF45, the inhibitor of DFF40 (the *DFFB* product). Like *DFFB*, *DFFA* has previously been mapped to chromosome 1p36.2-p36.3 (Leek et al. 1997). We were interested by this co-localization of *DFFA* and *DFFB*, especially since a dot matrix comparison of the DFFA/DFF45 and DFFB/DFF40 amino-acid sequences reveals regions of similarity (Fig. 1). The largest such region is at the N-terminus of both proteins, the end of this region (DFF40 residue 80) corresponding to the position of intron 2 in the *DFFB* gene. This suggested that, although the exon-intron structure of *DFFA* has yet to be defined, *DFFA* and *DFFB* may have evolved partly through local gene duplication.

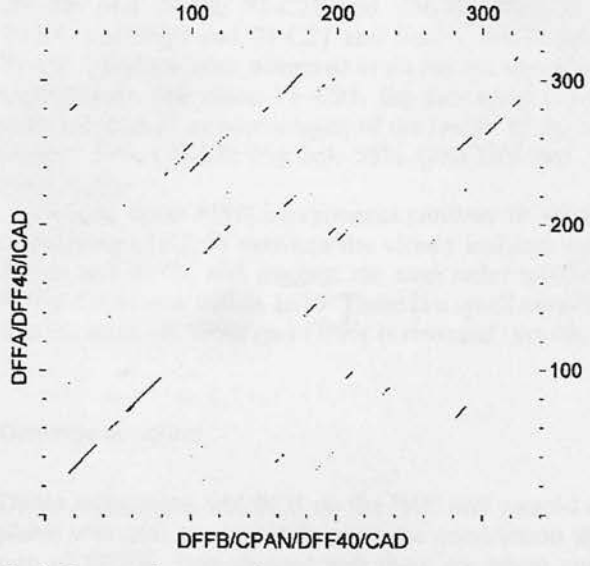


Fig. 1 Dot-matrix comparison between peptide sequences of DFF45 (*DFFA* gene product) and DFF40 (*DFFB* gene product). The comparison was performed using UWGCG COMPARE (window size 30, stringency 11)

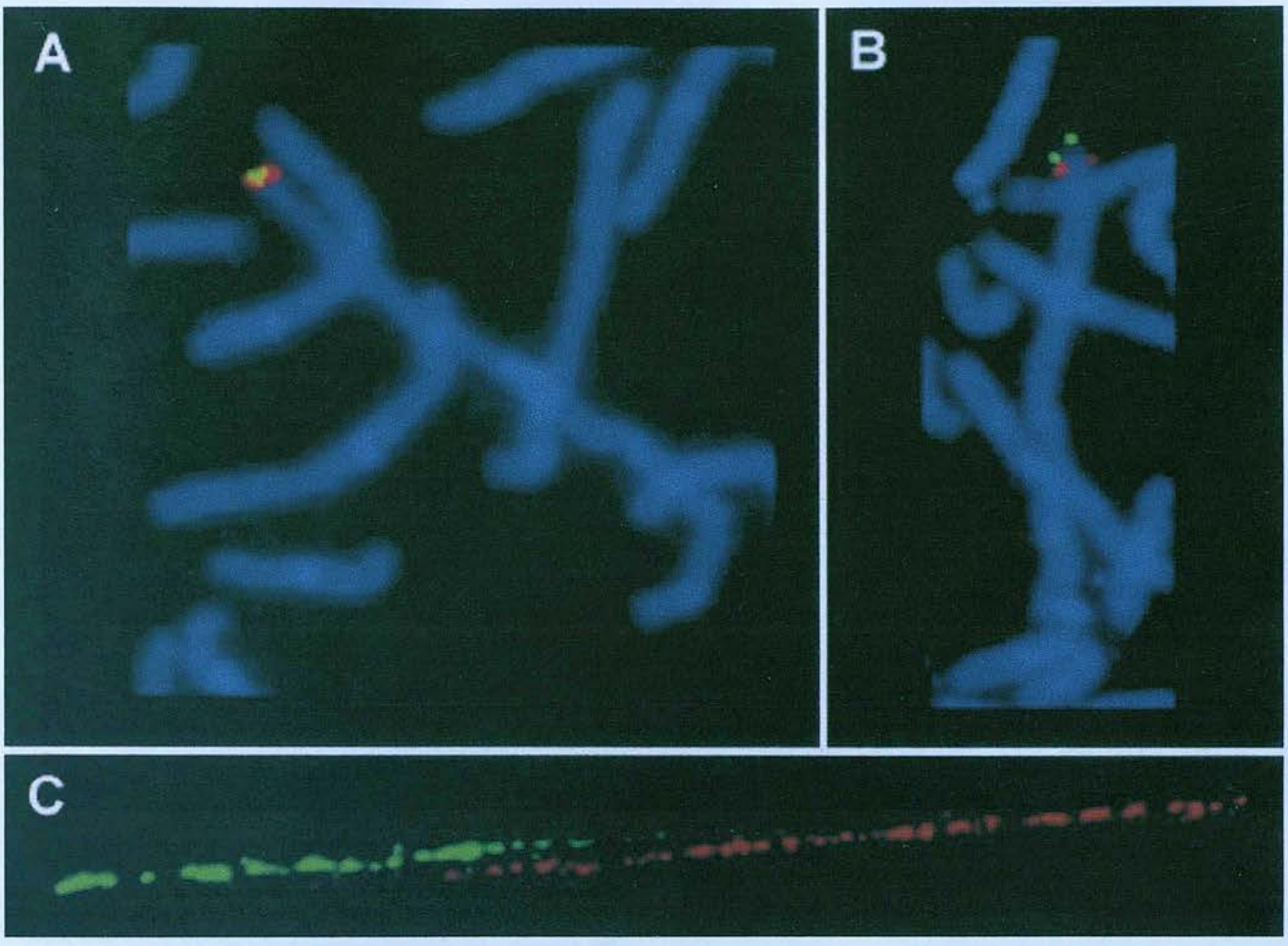


Fig. 2 Two-colour hybridisation on prometaphase chromosomes of **a** PAC clone 79–L8 (*TP73, green)* and cosmid clone cos176g8 (*DFFB, red*), **b** cosmid clone cos176g8 (*DFFB, green*) and PAC clone 92-K14 (*DFFA, red*), **c** fibre-FISH for PAC clone 79–L8 (*TP73, red)* and PAC clone 71-C21 (*green*)

To map the relative positions of *DFFA* and *DFFB,* dual colour FISH with PAC clone 92-K14 (containing *DFFA*; Leek et al. 1997) and cosmid 176g8 (*DFFB*; this study) was performed. On prometaphase chromosomes, these probes could be clearly separated, with *DFFB* (green) mapping distal to *DFFA* (red) (Fig. 2b). In contrast, as expected from their physical proximity, signals from the *DFFB*-containing cosmid and the *TP73* PAC clone 79-L8 overlapped, both on prometaphase chromosomes (Fig. 2a) and in interphase nuclei (not shown).

The gene for atrial natriuretic peptide precursor A (*NPPA*) is also known to map to this region of 1p36 (Arden et al. 1995; Yang-Feng et al. 1985). To define the map position of *DFFA* relative to *NPPA*, the *DFFA* PAC 92-K14 was co-hybridised with a cosmid probe for *NPPA*. In 13 metaphases, hybridisation signals for both probes overlapped, in 6 metaphases *NPPA* mapped distal and in one metaphase the inverse order was observed (not shown). Therefore, *DFFA* maps in the vicinity of (and most likely proximal to) *NPPA*.

The integrity and physical relationship between the clones in the *TP73-DFFB* contig was confirmed by fibre-FISH. Pair-wise hybridisations were performed for clones 286-D6 and 79-L8, 71-C21 and 286-D6, 71C-21 and 79-L8, cos176g8 and 71-C21 and finally, cos176g8 and 79-L8. Overlaps were observed in all but the last of these experiments. For clone 71–C21, the following overlaps were calculated, as percentages of the length of the other clones: 23% (79-L8; Fig. 2c), 55% (286-D6) and 19% (cos176g8).

Overall, these FISH experiments confirm the physical relationships (Fig. 3) between the clones isolated around *DFFB* and *TP73*, and suggest the map order tel-*DFFB-NPPA-DFFA*-cen within 1p36. There is a small possibility that the order of *NPPA* and *DFFA* is reversed (see above).

Genomic structure

Direct sequencing and PCR on the PAC and cosmid templates was next used to determine the exon-intron structure of *DFFB*. This showed that there are seven coding exons, spanning approximately 10 kb. This genomic structure is presented in Fig. 3. The genomic sequences adjacent to each splice junction are listed in Table 2.

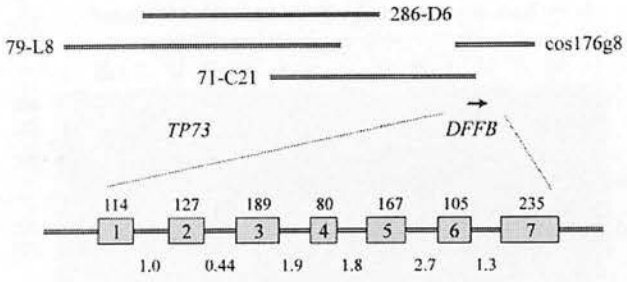


Fig. 3 The exon-intron organisation of the *DFFB* gene. The upper part shows the relationships between the clones used in this study, with the *small arrow* indicating the position and direction of transcription of *DFFB*. Below, the approximate intron sizes (in kilobases), and exact exon sizes (in base pairs) are given. For exons 1 and 7, the figure is for the size of the coding region, since the cap and poly(A) sites have not been defined

Attempts to amplify *DFFB* from genomic DNA using primers located in the 3' part of the DFF40 coding region (exons 6 and 7) yielded products differing in sequence from the published CPAN/DFF40 cDNA sequence (Halenbeck et al. 1998). These sequences proved however to be identical to those of a putative transcribed CAD pseudogene (Mukae et al. 1998). Since intron 6 was absent from our pseudogene-derived genomic DNA product, we conclude that this pseudogene is probably processed. A pseudogene-specific PCR assay was then designed, using new primers in exons 6 and 7. Under the conditions used, this PCR fails to amplify across intron 6, and generates only a 143-nt product specific for the pseudogene. In this way, the pseudogene was mapped to chromosome 9, using a single-chromosome somatic cell hybrid panel (Kelsell et al. 1995).

Mutation screening

The candidacy of *DFFB* as a neuroblastoma tumour suppressor gene was next addressed, through rigorous mutation screening, by complete DNA sequencing. A panel of 41 tumour or tumour cell line (Table 1) DNAs was used. Both 1p36-deleted and non-deleted tumours were analysed, since alternative hypotheses concerning the putative tumour suppressor gene are tenable. If the tumour suppressor gene were not imprinted, "second hit" mutations in the non-deleted allele might be expected, inviting analysis of 1p36-deleted tumours. On the other hand, if the tumour suppressor gene were imprinted, point mutations might not be needed as second hits in deleted cases, whereas non-deletion cases might have heterozygous point mutations.

We sequenced all seven *DFFB* exons, in all 41 DNA samples. No coding region mutations were detected in any of the samples. However, seven intronic and three coding region polymorphisms were detected (Table 3). Two of the three exonic polymorphisms are silent, while the third results in a conservative amino acid substitution of arginine for lysine 196. All three exonic polymorphisms were found in both tumour DNAs and normal controls.

Table 2 Sequences at the *DFFB* splice junctions. Exon sequences are *underlined*. Twenty bases of intron sequence are given at each side of each splice junction; for exons 1 and 7 the first and last 20 bases of the open reading frame are shown

Splice acceptor sequence	Exon	Splice donor sequence
ATGCTCCAGAAGCCCAAGAG	-1-	AGGGCTGTCTCCGCTTCCAGGTGCCCGCTGGGCTAGGCGG
TGCTTCTCCCGTCCCTGCAGCTCCCTGAGCGCGGTTCCCG	-2-	GGGCCAGGCCTGGCAGGGCTGTGAGTGGCAAGGACTTTGG
GTTTGTCCCATTGGTGGCAGATGTGAGCGACATCAGGCGC	-3-	GGACCCGCCGTGGTTTGAAGGTGCGTGGGGGCTGCAGCTG
TCTTCTCGTTTTCCTTGCAGGCTTGGAGTCCCGATTTCAG	-4-	TCCGGAGTTACCTGAGGGAGGTGAGCCTGAGTGAAGACCG
ACGTTCCTTGGTTCCTCCAGGTGAGCTCCTACCCCTCCAC	-5-	AAGGCTGGTTCTCCTGCCAGGTGAGCTGTGTGCCCTTTAT
CATTGTCTTTTGGCCCCCAGGGTCCCTTTGACATGGACAG	-6-	ACCTGGAACCTGGATCACATGTAAGCTCACAGAGCGAGGT
TCTCTAACCTTACTTTGCAGAATAGAAAAGAAACGCACCA	-7-	CTGTGCGGAAACGCCAGTGA

Table 3 Details of polymorphisms within the *DFFB* gene, identified during mutation screening. The numbering of nucleotides in the 5' untranslated region is relative to the first nucleotide of the ATG translation start codon (*ND* not determined, *nt* nucleotide, *UTR* untranslated region)

Position	Nature of polymorphism		Allele frequencies in normal controls	
Exon 5, nt 21	G→A	Silent	G=18/30	A=12/30
Exon 5, nt 77	G→A	Lys^{196}→Arg	G=27/30	A=3/30
Exon 7, nt 172	G→A	Silent	G=58/108	A=50/108
5' UTR, –38	A→G		ND	
5' UTR, –30	T→C		ND	
IVS 1, +15	A→G		ND	
IVS 1, +31	T→G		ND	
IVS 3, +44	T→C		ND	
IVS 3, –72	G→A		ND	
IVS 3, –82	T→C		ND	

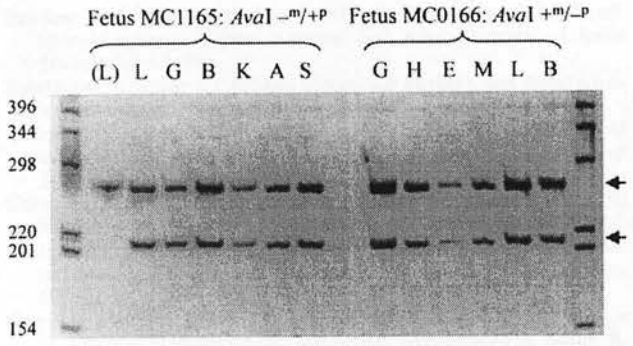


Fig. 4 Analysis of the allelic origin of *DFFB* transcripts. Ethidium bromide-stained polyacrylamide gel, showing *Ava*I-digested RT-PCR products spanning a polymorphic site at codon 318 in exon 7 of *DFFB*. Samples from two fetuses of opposite genotype are shown: *Ava*I $-^{m}/+^{p}$ (MC1165) and $+^{m}/-^{p}$ (MC0166). Undigested RT-PCR product is shown in the lane (*L*) adjacent to the marker. Marker band sizes are in bp. Predicted sizes for the undigested ('A' allele) and digested ('G' allele) products are 259 and 203 bp, respectively. (*L* lung, *G* gut, *B* brain, *K* kidney, *A* adrenal, *S* stomach, *H* heart, *E* eye, M muscle)

Imprinting

We have previously used a panel of first trimester fetal tissues to examine the allelic origin of transcripts of genes suspected to be imprinted (Campbell et al. 1994; Hayward et al. 1998). We genotyped 54 fetal DNA samples at the exon 7 *DFFB* polymorphism, which results in loss of an *Ava*I site if the 'A' allele is present. Twenty-eight fetuses were heterozygous, of which six had a homozygous mother and were therefore informative as to the origin of each allele (paternal samples were not available). To determine whether *DFFB* is imprinted, we chose one heterozygous fetus with a G/G mother (*Ava*I $+^{m}/-^{p}$), and one with an A/A mother (*Ava*I $-^{m}/+^{p}$). RNA samples from a range of tissues were available for each of these fetuses. RT-PCR was performed using primers in exons 6 and 7, followed by *Ava*I digestion to determine which alleles were represented in the RT-PCR products. The results indicated that both parental alleles were represented in the RNA samples, and it is therefore unlikely that *DFFB* is imprinted (Fig. 4).

Discussion

Various genes have been proposed as candidates for the putative tumour suppressor at 1p36.2–36.3 (Kaghad et al. 1997; Maris et al. 1997). However, none of these has yet been shown to carry coding region mutations, making it less likely that they play significant roles in the pathogenesis of neuroblastoma.

In this paper, we describe the structure of the newly identified gene *DFFB* that encodes the active subunit of DNA fragmentation factor (DFF40, CAD or CPAN). We have also assessed the candidacy of this mediator of apoptotic DNA fragmentation as the putative 1p36 tumour suppressor gene. Comprehensive mutation screening of the seven *DFFB* exons in a panel of 41 tumours and cell lines revealed several polymorphisms, but no tumour-specific mutations. As for other candidate genes lacking mutations, our results suggest that *DFFB* is unlikely to act as a tumour suppressor gene in neuroblastoma or Merkel cell carcinoma.

Despite this negative result, there are reasons for not completely discounting the potential role of *DFFB* in neoplasia. Firstly, the fact that *DFFB* maps to a chromosomal region believed to be subject to genomic imprinting raises the question of whether *DFFB* itself could be imprinted. This possibility is not completely excluded by our RT-PCR experiments. If *DFFB* were an imprinted gene, expressed solely from one parental allele, consistent deletion or inactivation of the active allele could result in loss of DFF40 activity without the presence of a "second-hit" mutation in the remaining allele. This possibility, however, is made less likely by the fact that we have found no evidence for *DFFB* mutations in non-deleted neuroblastomas, in which two hits might be expected to be required.

Secondly, the investigation of *DFFB* in other tumours may be worthwhile. Several tumours other than neuroblastoma have been associated with genetic alterations in the 1p36 region, including dysplastic naevus and familial melanoma (Ping et al. 1998), breast cancer (Bieche et al. 1998; Nagai et al. 1995), parathyroid adenoma (Tahara et al. 1997), ovarian cancer (Thompson et al. 1997), non-small-cell lung cancer (Gasparian et al. 1998), and renal oncocytoma (Thrash-Bingham et al. 1996).

Finally, the failure to find point mutations within individual candidate tumour suppressor genes within 1p36 may still allow the possibility that simultaneous haploinsufficiency for several clustered genes (such as *TP73* and *DFFB*) with roles in cell cycle control, differentiation and apoptosis could be a step in the process of transformation. This suggests that other functional assays for investigating the role of genes such as *DFFB* may need to be considered.

Acknowledgements Human genomic library reagents and somatic cell hybrid DNAs were obtained from the UK HGMP Resource Centre. H.J. holds an MRC postgraduate studentship. N.V.R. is a postdoctoral researcher of the Fund for Scientific Research, Flanders. We thank the following for supplying tumour material or cell lines: Dr. P. Ambros, Dr. J.H. Leonard, Dr. G. Laureys, Dr. R. Sciot, Prof. Dr. J.M. Naeyaert, Prof. Dr. S. Van Belle, Dr. J. Bridge, Dr. S.G. Ronan, Dr. S.T. Rosen. This study was supported by grants from the Medical Research Council (U.K.), Flemish Institute for the Promotion of Scientific Technological Research in Industry (IWT), GOA grant number 12051397 and FWO grant 3.0085.96 N.

References

Arden KC, Viars CS, Weiss S, Argentin S, Nemer M (1995) Localization of the human B-type natriuretic peptide precursor (NPPB) gene to chromosome 1p36. Genomics 26:385–389

Bieche I, Khodja A, Lidereau R (1998) Deletion mapping in breast tumor cell lines points to two distinct tumor-suppressor genes in the 1p32-pter region, one of deleted regions (1p36.2) being located within the consensus region of LOH in neuroblastoma. Oncol Res 5:267–272

Brodeur GM (1995) Genetics of embryonal tumours of childhood: retinoblastoma, Wilms' tumour and neuroblastoma. Cancer Surveys 25:67–99

Brodeur GM, Fong CT (1989) Molecular biology and genetics of human neuroblastoma. Cancer Genet Cytogenet 41:153–174

Campbell R, Gosden CM, Bonthron DT (1994) Parental origin of transcription from the human GNAS1 gene. J Med Genet 31:607–614

Caron H, Sluis P van, Hoeve M van, Kraker J de, Bras J, Slater R, Mannens M, Voute PA, Westerveld A, Versteeg R (1993) Allelic loss of chromosome 1p36 in neuroblastoma is of preferential maternal origin and correlates with N-myc amplification. Nat Genet 4:187–190

Caron H, Peter M, Sluis P van, Speleman F, Kraker J de, Laureys G, Michon J, Brugieres L, Voute PA, Westerveld A, Slater R, Delattre O, Versteeg R (1995) Evidence for two tumour suppressor loci on chromosomal bands 1p35–36 involved in neuroblastoma: one probably imprinted, another associated with N-myc amplification. Hum Mol Genet 4:535–539

Caron H, Sluis P van, Kraker J de, Bokkerink J, Egeler M, Laureys G, Slater R, Westerveld A, Voute PA, Versteeg R (1996) Allelic loss of chromosome 1p as a predictor of unfavourable outcome in patients with neuroblastoma. N Engl J Med 334:225–230

Cheng NC, Van Roy N, Chan A, Beitsma M, Westerveld A, Speleman F, Versteeg R (1995) Deletion mapping in neuroblastoma cell lines suggests two distinct tumour suppressor genes in the 1p35–36 region, only one of which is associated with N-myc amplification. Oncogene 10:291–297

Enari M, Sakahira H, Yokoyama H, Okawa K, Iwamatsu A, Nagata S (1998) A caspase-activated DNase that degrades DNA during apoptosis, and its inhibitor ICAD. Nature 391:43–50

Gasparian AV, Laktionov KK, Belialova MS, Pirogova NA, Tatosyan AG, Zborovskaya IB (1998) Allelic imbalance and instability of microsatellite loci on chromosome 1p in human non-small-cell lung cancer. Br J Cancer 77: 1604–1611

Gingrich JC, Boehrer D, Garnes JA, Johnson W, Wong B, Bergmann A, Eveleth GG, Longlois RG, Carrano AV (1996) Construction and characterisation of human chromosome 2-specific cosmid, fosmid, and PAC clone libraries. Genomics 32:65–74

Halenbeck R, MacDonald H, Roulston A, Chen TT, Conroy L, Williams LT (1998) CPAN, a human nuclease regulated by the caspase-sensitive inhibitor DFF45. Curr Biol 8:537–540

Hayward BE, Kamiya M, Strain L, Moran V, Campbell R, Hayashizaki Y, Bonthron DT (1998) The human *GNAS1* gene is imprinted and encodes distinct paternally and biallelically expressed G proteins. Proc Natl Acad Sci USA 95:10038–10043

Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, Jong PJ de (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. Nat Genet 6:84–89

Jost CA, Marin MC, Kaelin WG (1997) p73 is a human p53-related protein that can induce apoptosis. Nature 389:191–194

Kaghad M, Bonnet H, Yang A, Creancier L, Biscan JC, Valent A, Minty A, Chalon P, Lelias JM, Dumon X, Ferrara P, McKeon F, Caput D (1997) Monoallelically expressed gene related to p53 at 1p36, a region frequently deleted in neuroblastoma and other human cancers. Cell 90:809–819

Kelsell DP, Rooke L, Warne D, Bouzyk M, Cullin L, Cox S, West L, Povey S, Spurr NK (1995) Development of a panel of monochromosomal somatic cell hybrids for rapid gene mapping. Ann Hum Genet 59:233–241

Kovalev S, Marchenko N, Swendeman S, LaQuaglia M, Moll UM (1998) Expression level, allelic origin, and mutation analysis of the p73 gene in neuroblastoma tumors and cell lines. Cell Growth Differ 9:897–903

Kuida K, Zheng TS, Na S, Kuan C, Yang D, Karasuyama H, Rakic P, Flavell RA (1996) Decreased apoptosis in the brain and premature lethality in CPP32-deficient mice. Nature 384:368–372

Leek JP, Carr IM, Bell SM, Markham AF, Lench NJ (1997) Assignment of the DNA fragmentation factor gene (DFFA) to human chromosome bands 1p36.3–>p36.2 by in situ hybridisation. Cytogenet Cell Genet 79:212–213

Liu X, Zou H, Slaughter C, Wang X (1997) DFF, a heterodimeric protein that functions downstream of caspase-3 to trigger DNA fragmentation during apoptosis. Cell 89:175–184

Mai M, Qian C, Yokomizo A, Tindall DJ, Bostwick D, Polychronakos C, Smith DI, Liu W (1998) Loss of imprinting and allele switching of p73 in renal cell carcinoma. Oncogene 17:1739–1741

Maris JM, White PS, Beltinger CP, Sulman EP, Castleberry RP, Shuster JJ, Look AT, Brodeur GM (1995) Significance of chromosome 1p loss of heterozygosity in neuroblastoma. Cancer Res 55:4664–4669

Maris JM, Jensen J, Sulman EP, Beltinger CP, Allen C, Biegel JA, Brodeur GM, White PS (1997) Human Kruppel-related 3 (HKR3): a candidate for the 1p36 neuroblastoma tumour suppressor gene? Eur J Cancer 33:1991–1996

Mukae N, Enari M, Sakahira H, Fukuda Y, Inazawa J, Toh H, Nagata S (1998) Molecular cloning and characterisation of human caspase-activated DNase. Proc Natl Acad Sci USA 95:9123–9128

Nagai H, Negrini M, Carter SL, Gillum DR, Rosenberg AL, Schwartz GF, Croce CM (1995) Detection and cloning of a common region of loss of heterozygosity at chromosome 1p in breast cancer. Cancer Res 55:1752–1757

Nomoto S, Haruki N, Kondo M, Konishi H, Takahashi T (1998) Search for mutations and examination of allelic expression imbalance of the *p73* gene at 1p36.33 in human lung cancers. Cancer Res 58:1380–1383

Ping YJ, Nakatsu Y, Goldstein AM, Tucker MA, Kraemer KH, Tanaka K (1998) RPA2, a gene for the 32 kDa subunit of replication protein A on chromosome 1p35–36, is not mutated in patients with familial melanoma linked to chromosome 1p36. Melanoma Res 8:47–52

Reik W, Maher ER (1997) Imprinting in clusters: lessons from Beckwith-Wiedemann syndrome. Trends Genet 13:330–334

Sakahira H, Enari M, Nagata S (1998) Cleavage of CAD inhibitor in CAD activation and DNA degradation during apoptosis. Nature 391:96–99

Schwab M, Praml C, Amler LC (1996) Genomic instability in 1p and human malignancies. Cancer Genet Cytogenet 16:211–229

Speleman F, Van Gele M, Maertens L, Van Roy N (1997) Improved protocol for the preparation of chromatin fibres from fixed cells. Technical Tips Online (URL: http://www.elsevier.com/locate/tto), T01123

Tahara H, Smith AP, Gaz RD, Zariwala M, Xiong Y, Arnold A (1997) Parathyroid tumor suppressor on 1p: analysis of the p18 cyclin-dependent kinase inhibitor gene as a candidate. J Bone Miner Res 12:1330–1334

Takita J, Hayashi Y, Yokota J (1997) Loss of heterozygosity in neuroblastomas – an overview. Eur J Cancer 33:1971–1973

Thompson FH, Taetle R, Trent JM, Liu Y, Massey-Brown K, Scott KM, Weinstein RS, Emerson JC, Alberts DS, Nelson MA (1997) Band 1p36 abnormalities and t(1;17) in ovarian carcinoma. Cancer Genet Cytogenet 96:106–110

Thrash-Bingham CA, Salazar H, Greenberg RE, Tartof KD (1996) Loss of heterozygosity studies indicate that chromosome arm 1p harbors a tumor suppressor gene for renal oncocytomas. Genes Chromosomes Cancer 16:64–67

Trask BJ, Engh G van den, Christensen M, Massa HF, Gray JW, Van Dilla MA (1991) Characterisation of somatic cell hybrids by bivariate flow karyotyping and fluorescence in situ hybridisation. Somat Cell Molec Genet 17:117–136

Van Gele M, Speleman F, Vandesompele J, Van Roy N, Leonard JH (1998a) Characteristic pattern of chromosomal gains and losses in Merkel cell carcinoma detected by comparative genomic hybridisation. Cancer Res 58:1503–1508

Van Gele M, Van Roy N, Ronan SG, Messiaen L, Vandesompele M J, Geerts ML, Naeyaert JM, Blennow E, Bar-Am I, Das Gupta TK, Drift P van der, Versteeg R, Leonard JH, Speleman F (1998b) Molecular analysis of 1p36 breakpoints in two Merkel cell carcinomas. Genes Chromosomes Cancer 23:67–71

Van Roy N, Laureys G, Cheng NC, Willem P, Opdenakker G, Versteeg R, Speleman F (1994) 1;17 translocations and other chromosome 17 rearrangements in human primary neuroblastoma tumors and cell lines. Genes Chromosomes Cancer 10: 103–114

Vandesompele J, Van Roy N, Van Gele M, Laureys G, Ambros P, Heimann P, Devalck C, Schuuring E, Brock P, Otten J, Gyselinck J, DePaepe A, Speleman F (1998) Genetic heterogeneity of neuroblastoma studied by comparative genomic hybridisation. Genes Chromosomes Cancer 23:141–152

Yang-Feng TL, Floyd-Smith G, Nemer M, Drouin J, Francke U (1985) The pronatriodilatin gene is located on the distal short arm of human chromosome 1 and on mouse chromosome 4. Am J Hum Genet 37:1117–1128

Human Molecular Genetics, 2000, Vol. 9, No. 3 453–460

The cell cycle control gene *ZAC/PLAGL1* is imprinted—a strong candidate gene for transient neonatal diabetes

Mamoru Kamiya[1,2], Hannah Judson[3], Yasushi Okazaki[1], Moriaki Kusakabe[1], Masami Muramatsu[1], Shuji Takada[1,4], Nobuo Takagi[4], Takahiro Arima[1,5], Norio Wake[5], Katsunori Kamimura[6], Kenichi Satomura[7], Robert Hermann[8], David T. Bonthron[3] and Yoshihide Hayashizaki[1,2,+]

[1]CREST, Japan Science and Technology Corporation (JST), Genome Exploration Research Group, Genomic Sciences Center (GSC), Genome Science Laboratory and Biogenetic Research Center, Riken Tsukuba Life Science Center, the Institute of Physical and Chemical Research (RIKEN), Ibaraki 305-0074, Japan, [2]Institute of Basic Medical Sciences, University of Tsukuba, Ibaraki 305-0006, Japan, [3]Molecular Medicine Unit, University of Leeds, Leeds LS9 7TF, UK, [4]Graduate School of Environmental Earth Science, Hokkaido University, Hokkaido 060-0808, Japan, [5]Department of Reproductive Physiology and Endocrinology, Medical Institute of Bioregulation, Kyushu University, Oita 874-0838, Japan, [6]Department of Pediatrics, Kobe City General Hospital, Kobe 650-0046, Japan, [7]Department of Pediatrics, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka 594-1101, Japan, [8]Department of Pediatrics, University Medical School of Pécs, Pécs 7623, Hungary

Received 15 November 1999; Revised and Accepted 7 December 1999

We describe a screen for new imprinted human genes, and the identification in this way of *ZAC* (zinc finger protein which regulates apoptosis and cell cycle arrest)/*PLAGL1* (pleomorphic adenoma of the salivary gland gene like 1) as a strong candidate gene for transient neonatal diabetes mellitus (TNDM). To screen for imprinted genes, we compared parthenogenetic DNA from the chimeric patient FD and androgenetic DNA from hydatidiform mole, using restriction landmark genome scanning for methylation. This resulted in identification of two novel imprinted loci, one of which (NV149) we mapped to the TNDM region of 6q24. From analysis of the corresponding genomic region, it was determined that NV149 lies ~60 kb upstream of the *ZAC/PLAGL1* gene. RT–PCR analysis was used to confirm that this *ZAC/PLAGL1* is expressed only from the paternal allele in a variety of tissues. TNDM is known to result from upregulation of a paternally expressed gene on chromosome 6q24. The paternal expression, map position and known biological properties of *ZAC/PLAGL1* make it highly likely that it is the TNDM gene. In particular, *ZAC/PLAGL1* is a transcriptional regulator of the type 1 receptor for pituitary adenylate cyclase-activating polypeptide, which is the most potent known insulin secretagog and an important mediator of autocrine control of insulin secretion in the pancreatic islet.

INTRODUCTION

Genomic imprinting refers to differential contributions from the two parental genomes, especially in the development and/or differentiation of the mammalian embryo. Its existence was originally demonstrated by nuclear transplantation experiments (1,2). The non-equivalence of the parental genomes has been explained by the differential expression of genes from one or other allele (1,3). Approximately 35 mouse (4) and 40 human genes regulated by imprinting have been reported, many of which, however, have not been fully characterized (5). It has become clear that imprinting is involved in a number of human genetic disorders, such as Beckwith–Wiedemann syndrome, Wilms' tumor, Prader–Willi syndrome and Angelman syndrome (5). One approach to identifying the genes involved in imprinted diseases is to screen systematically for and map imprinted loci. Strong candidate genes may be more readily identified through this approach than through traditional positional cloning, because such candidate genes will by definition satisfy both criteria of correct chromosomal location and imprinted expression pattern.

In order to achieve this goal, systematic screening for human imprinted genes is required. Several methods for systematically searching for imprinted genes within the mouse genome have been reported (6). The methylation-based screening method, restriction landmark genomic scanning with methylation-sensitive restriction endonuclease (RLGS-M), was developed by our group (7). The mouse strategy relies on comparing the polymorphic RLGS-M patterns derived from reciprocal crosses between inbred parental strains. Another approach is to rely on differential mRNA expression; differential display or subtraction cloning using mRNA from parthenogenetic and

+To whom correspondence should be addressed. Tel: +81 298 36 9145; Fax: +81 298 36 9098; Email: yosihide@rtc.riken.go.jp

androgenetic embryos is therefore also a successful method for identifying imprinted genes (6).

In contrast, screening for human imprinted genes has several technical and ethical limitations in terms of the availability of the samples. RLGS-M relies on polymorphic differences between inbred strains. The required reciprocal crosses are not available in humans. As far as expression cloning is concerned, both the differential display and subtractive cDNA cloning approaches require uniparental RNA samples prepared by embryo manipulation, and therefore again are not applicable to human material. Furthermore, those human genes which are orthologs of mouse imprinted genes are not always imprinted (8). Thus, a new method for the screening for human imprinted genes is required, which relies neither on inheritance of DNA polymorphisms nor on the study of the human orthologs of mouse imprinted genes.

One reagent which has proven valuable in systematic screening for human imprinted genes is DNA from the parthenogenetic chimera FD. This was the first reported patient with chimerism between normal (biparental) and parthenogenetic cells (9). The leukocytes, in particular, were found to be 100% parthenogenetic. For comparison with the FD sample, complete hydatidiform mole is also useful; this tumor is derived from androgenetic extraembryonic tissue (10). An mRNA-based comparison between these two tissues is not appropriate, since: (i) the amount of the available material from the patient is limited; and (ii) only a limited number of genes are expressed in these specific tissues. On the other hand, DNA methylation appears to be an invariable accompaniment to genomic imprinting, and a screening system based on methylation is therefore attractive, and not restricted to the subset of genes expressed in any one tissue. We have therefore analyzed DNAs from the above-mentioned human uniparental sources by RLGS-M. This method employs ^{32}P-end-labeling of rare-cutting restriction sites in genomic DNA, followed by high resolution two-dimensional electrophoresis separation of the fragments (11). It permits analysis of >2500 loci on one autoradiograph. It also preferentially targets transcription units because of its dependence on cleavage at rare CpG-containing sites. In addition, RLGS-M is dosage-sensitive, and can distinguish haploid and diploid intensity of single-copy genes. This important property confers the ability to screen for imprinted genes by comparing spot intensity in parthenogenetic and androgenetic DNA, without having to rely on polymorphisms to distinguish the parental alleles.

Transient neonatal diabetes mellitus (TNDM) is one genetic disease that shows an imprinting effect; it appears to result from upregulation of a paternally expressed gene in 6q24, since this phenotype can result from paternal uniparental disomy of chromosome 6 and from duplications of this chromosomal region (12–14). TNDM patients have insulin-dependent diabetes lasting for the first few months of life. In many cases, the diabetes subsequently returns permanently in mid-childhood. In this paper, we report the results of a methylation-based RLGS screen for human imprinted genes, using parthenogenetic and androgenetic DNAs, and the principle that differential methylation 'tags' an imprinted locus (7,15). In this screen, we identified two imprinted genes: *GNAS1*, located on human chromosome 20q13.2 (16), and, as reported here, *ZAC* (zinc finger protein which regulates apoptosis and cell cycle arrest). *ZAC* is known to be capable of inducing G_1 cell-cycle arrest and apoptosis, and to be a transcriptional regulator of the type 1 receptor for the pituitary adenylate cyclase activating peptide ($PACAP_1$-R) (17,18). PACAP is an important intra-islet regulator of insulin secretion (19,20), and this, together with the finding that the imprinted *ZAC* gene maps to the TNDM region of 6q24, makes *ZAC* a compelling TNDM candidate gene.

RESULTS

RLGS-M screening for imprinted genes using FD and hydatidiform mole DNA

RLGS is a multiplex genome scanning system allowing visualization of >2500 spots/loci, based on the principle that restriction enzyme sites can be used as landmarks. The technology employs direct labeling of genomic DNA and high resolution two-dimensional DNA electrophoresis (11,21). The use of methylation-sensitive enzymes enables detection of the methylation status of the restriction landmark (22). A methylated landmark is not cleaved by the landmark enzyme and yields no spot signals whereas, if the landmark is unmethylated, the cleaved end of the restriction fragment is labeled to produce spot signals. Haploid and diploid copy number of the unmethylated restriction landmark give half and full intensity of the autoradiographic spots, respectively (23).

In higher organisms, the 5′ position of cytosine residues is the only site for DNA methylation, the bulk of which occurs at the dinucleotide CpG. The total number of 'CpG islands', often located upstream of genes, is estimated to be ~30 000 (24). Methylation of cytosines within a CpG island often reflects transcriptional silencing of the gene (22). Rare cutting restriction enzymes with GC-rich recognition sites cleave preferentially within CpG islands. For example, 90% of total *Not*I sites (GCGGCCGC) are located in CpG islands, which in turn implies that 90% of RLGS spots produced using *Not*I have the potential to reflect the transcriptional status of a CpG island-associated gene (7,11).

Leukocytes from the patient FD are 100% parthenogenetic and isodisomic (9). Conversely, complete hydatidiform mole is derived from androgenetic extraembryonic tissue (25). Although in previous applications of RLGS (7,15,26), the segregation of polymorphic spots was used to provide information about linkage or imprinting, a comparison between parthenogenetic and androgenetic materials allows the use of all non-polymorphic spots, providing a very powerful screening tool.

The enzyme combinations of *Not*I–*Pvu*II–*Pst*I, *Not*I–*Pst*I–*Pvu*II and *Asc*I–*Eco*RV–*Mbo*I were used for the screening. Because of the uniform isodisomy of all 23 chromosome pairs, the spot intensities on RLGS-M of FD blood DNA are very uniform (Fig. 1a). The spot intensities of androgenetic tissue from hydatidiform mole were not as uniform, due to the additional methylation changes occurring during tumorigenesis (data not shown). A schematic figure of the strategy used to identify imprinted spots/loci is shown in Figure 1b. In this figure, loci A and C are maternally and paternally methylated, respectively, whereas B is a non-imprinted locus, unmethylated on both alleles. Since FD leucocytes have two maternally inherited alleles, maternally methylated locus A is not cleaved with *Not*I, resulting in the disappearance of the corresponding spot (Fig. 1b, FD). On the other hand, this spot shows diploid intensity in the complete mole (Fig. 1b, mole), corresponding to the presence only of unmethylated paternal alleles. In the normal biparental DNA, such a spot shows half intensity (Fig.

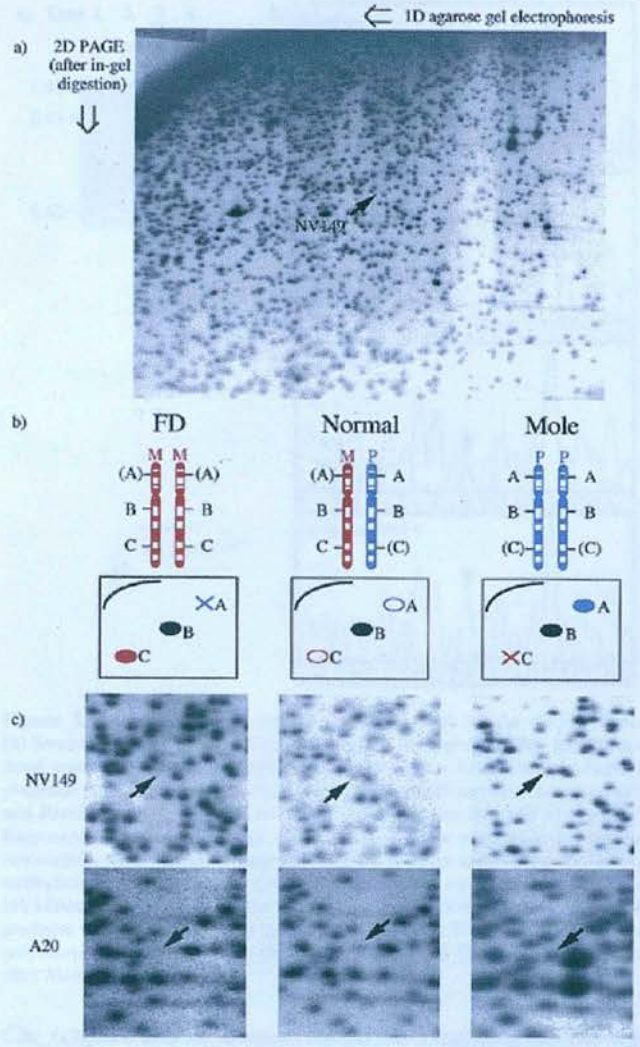


Figure 1. The two-dimensional profile of restriction landmark genomic scanning using methylation-sensitive restriction endonuclease (RLGS-M). (**a**) RLGS-M profile of FD blood DNA produced by the enzyme combination of *Not*I–*Pst*I–*Pvu*II. The arrow indicates the imprinted spot, *NV149*. (**b**) The principle of screening for imprinted spots. Spots A and C represent maternally and paternally methylated imprinted loci, respectively. Spot B is from a normal unmethylated locus. The imprinted spots show half intensity (haploid; open circle) and the non-imprinted unmethylated spot shows full intensity (diploid; closed circle). The maternally methylated spot (A) disappears in the parthenogenetic sample, FD, but this spot shows diploid intensity in androgenetic hydatidiform mole. Conversely, a spot which is paternally methylated appears in parthenogenetic DNA, and is absent from androgenetic DNA. (**c**) Close-up of RLGS-M profiles of *NV149* and *A20*. The *A20* RLGS-M profiles were produced by enzyme combination *Asc*I–*Eco*RV–*Mbo*I. Since these two spots were both maternally methylated, their intensities were zero, half and full in FD, normal and mole, respectively.

1b, normal). Conversely, the spot representing a paternally methylated locus C shows diploid, haploid and null intensity in RLGS-M profiles of FD, normal and mole, respectively.

Using these criteria, two spots named *NV149* and *A20* were identified as candidate imprinted loci. These are shown in Figure 1c. *NV149* was detected using the enzyme combination of *Not*I–*Pst*I–*Pvu*II and *A20* using *Asc*I–*Eco*RV–*Mbo*I (Table 1). These two spots were excised from the gel and cloned. The

Table 1. RLGS-M spots and related diseases

Spot name	Enzyme combination	Locus	Gene	Related disease
NV149	*Not*I–*Pst*I–*Pvu*II	6q24	*ZAC*	TNDM
A20	*Asc*I–*Eco*RV–*Mbo*I	20q13.2	*GNAS1*	PHP-IA

results of subsequent analysis of *A20*, which is located upstream of the imprinted *GNAS1* locus, have been previously reported (16). In this paper, we focus on the further analysis of the *NV149* clone.

Maternal allele-specific methylation of the *NV149* locus

The methylation status of *NV149* locus was analyzed using Southern blotting and sequencing (Fig. 2). The methylation status of the *Not*I, *Bss*HII and *Sma*I sites, located within the DNA fragment of the *NV149* spot clone (Fig. 3a), was analyzed by Southern blotting (Fig. 2a). Genomic DNA digested by *Pvu*II yields a 1.18 kb band (Fig. 2a, lane 1). About half of the molecules of this fragment were cleaved by each of the three methylation-sensitive restriction enzymes, confirming the likelihood of monoallelic methylation. To confirm the parental specificity of this monoallelic methylation (i.e. imprinted methylation), PCR direct sequencing was performed using DNA from a normal individual and his parent. In this assay (Fig. 2b), a DNA fragment spanning both the *Not*I site and a nearby single-nucleotide polymorphism was amplified, using primers NV149um1 and NV149l1 (Fig. 3a). The NV149u1 primer was used for sequencing (Fig. 3a). The paternal (A) and maternal (G) alleles at nucleotide 167 can be identified from the genotypes of the homozygous parents. Their progeny showed a heterozygous pattern of A/G (Fig. 3c). However, *Not*I digestion of the genomic DNA prior to PCR amplification eliminated the sequence signal from the paternal A allele. These data indicate that only the maternal allele is methylated, and that the methylation pattern of the genomic DNA at *NV149* is imprinted in the manner predicted from the RLGS screen.

NV149 maps to chromosome 6q24

As the next step, the spot clone *NV149* was mapped using a radiation hybrid panel and a gridded CEPH megaYAC library to human chromosome 6q24 (data not shown). Further analysis and library screening showed that *NV149* lies within PAC contig Chr_6ctg224 (http://webace.sanger.ac.uk/HGP/Chr6), as shown in Figure 3. In particular, PAC clone dJ340H11 (GenBank accession no. AL109755) contains the *NV149* spot clone sequence. PAC dJ340H11 also contains the whole of the gene encoding the C_2H_2-type zinc finger protein, *ZAC*. *ZAC* was originally reported as a human gene orthologous to mouse *Zac1*, which is a regulator of the type 1 receptor for PACAP and also an inducer of apoptosis and cell cycle arrest (17). PACAP is also known as an islet neuropeptide involved in the autocrine stimulation of insulin secretion (19,20).

ZAC/PLAGL1 expression profile

Three variants of the human *ZAC/PLAGL1* transcripts have been reported, named *ZAC* (27), *PLAGL1* (28) and *LOT1* (lost on transformation 1) (29). The genomic structure of the *ZAC/PLAGL1* region is shown in Figure 3c. From further analysis, we concluded that two alternative promoters, of *ZAC* and *PLAGL1*, are located in this region, but that *LOT1* is a chimerism artifact. The 5′-untranslated region (5′-UTR) of *LOT1* is not present in PAC contig

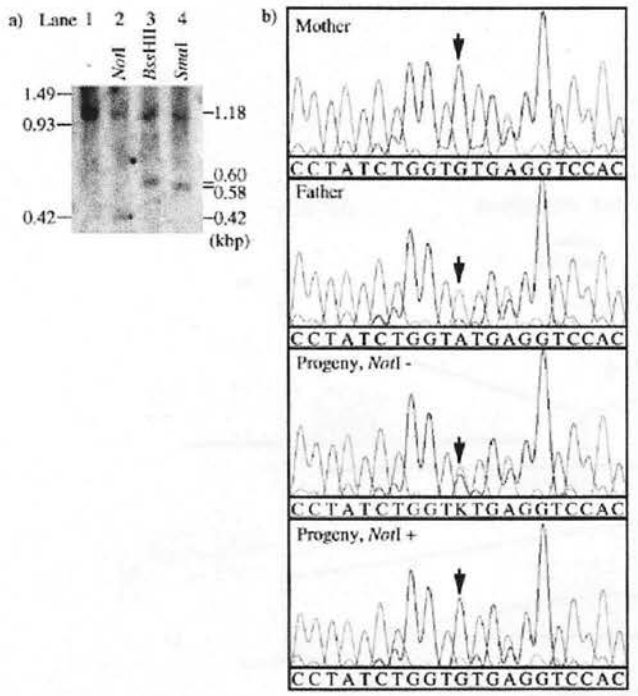


Figure 2. Imprinted methylation of genomic DNA in the *NV149* locus. (**a**) Southern hybridization using *NV149* clone as a probe. *Not*I, *Bss*HII and *Sma*I were used to test the monoallelic methylation in this region. Normal placental DNA was digested with *Pvu*II (lane 1), *Pvu*II and *Not*I (lane 2), *Pvu*II and *Bss*HII (lane 3) or *Pvu*II and *Sma*I (lane 4). Note that half of the *Pvu*II fragments were cleaved by each methylation-sensitive enzyme, indicating the monoallelic methylation of these sites. (**b**) PCR direct sequencing analysis of methylation imprinting. The genomic DNAs were amplified with primers, NV149um1 and NV149l1, after treatment with (+) or without (–) *Not*I. PCR products were sequenced with the primer, NV149u1. The arrow indicates the polymorphic (G/A) site. Only the maternally derived G allele was amplified after *Not*I digestion.

Chr_6ctg224, and on radiation hybrid (RH) mapping analysis was found to be on chromosome 3. This *LOT1* 5′-UTR sequence also completely matches part of the sequence of clone RPCI4-736H12 (GenBank accession no. AC006060) which again has been mapped to chromosome 3. We therefore conclude that the *LOT1* gene is a chimeric clone, carrying sequences derived from chromosome 3 and from the *ZAC* coding region on chromosome 6. We therefore excluded *LOT1* from further analysis. *PLAGL1* was identified by sequence similarity to *PLAG* (pleomorphic adenoma of the salivary gland gene) which is transcriptionally activated in these salivary gland tumors as a result of promoter swapping caused by the chromosome translocation, t(3;8)(p21;q12) (30). As shown in Figure 3c, *PLAGL1* and *ZAC* have different 5′ ends, but utilize common downstream exons.

To test the imprinted expression of *ZAC*, we identified polymorphisms within the *ZAC* exons. The first is a G/T substitution located at nucleotide 219 (*ZAC* exon 1). Direct sequencing across this site was performed on genomic PCR products and RT–PCR products derived from placental mRNA. The results are shown in Figure 4a. Maternal DNA has G at this position, whereas the father is a G/T heterozygote. Their daughter is also heterozygous. Her placental RT–PCR product was derived exclusively from the paternal T allele, showing that the expression of *ZAC* is monoallelic and paternal. Using a further G/T polymorphism at nucleotide 875, expression was subsequently examined in other tissues. Only 4 of 53 human fetuses were heterozygous at nucleotide 875. Monoallelic expression was seen in fetal heart, kidney, muscle, cord, adrenal gland and lung. However, *ZAC* is unfortunately biallelically expressed in white blood cells, precluding the analysis of TNDM patients' blood for evidence of relaxation of imprinting.

Both of the above polymorphisms lie within the *ZAC*-specific portion of *ZAC* exon 1, and are not found within the *PLAGL1* splice form. In order to examine *PLAGL1* expression, another polymorphism (C/G) in exon 2 (at nucleotide 1029 of *ZAC*) was employed. This polymorphism is detectable by restriction digestion with *Ava*II, which cuts the G allele (GGWCC), but not the C allele (GCWCC). RT–PCR primers PLAGL1ex1b+2f and PLAG1ex2r2 were designed to flank this polymorphism and be cDNA specific. They generate a 304 bp cDNA-specific product from both *ZAC1* and *PLAGL1* mRNA species. *Ava*II digestion products are 148, 130 and 26 bp for a nucleotide 1029 C allele, the G allele splitting the 148 bp fragment into 103 and 45 bp. For PCR on genomic DNA, the alternative upstream primer PLAGL1ex2f1 was used, yielding a 90 bp band in place of the 26 bp cDNA-specific band. The result, using placental mRNA and DNA from two individuals, is shown in Figure 4b. Both individuals' DNA is heterozygous at nucleotide 1029 (103 + 45 bp from the C allele and 148 bp from the G allele). Individual 1 is C^{mat}/G^{pat} and individual 2 G^{mat}/C^{pat} (data not shown). The RT–PCR product was derived exclusively from the paternal allele in both individuals (Fig. 4b, lanes 1r and 2r). Since this product represents a mixture of *PLAGL1* and *ZAC* transcripts, it suggests that, at least in placenta, both the *ZAC* and *PLAG1* promoters are active only on a paternal allele.

DISCUSSION

ZAC/PLAGL1 is a strong candidate gene for TNDM

As described above, *ZAC/PLAGL1* is located at 6q24, in the same chromosomal region as the imprinted TNDM locus (Fig. 5) (14). TNDM was mapped to an 18.72 cR region between *D6S1699* and *D6S1010*. The distal flanking marker, *D6S1010*, was defined by the extent of the region trisomic in an insertional translocation case [46XX, der(2) ins(2;6)(2pter-2p22.2::6q?22.32–?6q?23.1::2p22.2–2qter)]. The proximal flanking marker, *D6S1699*, was defined by linkage analysis in a family with inherited TNDM. The TNDM candidate interval between *D6S1699* and *D6S1010* is 18.72 cR, equivalent to <5.4 Mb. *ZAC/PLAGL1*, in comparison, maps 12.05 cR distal to *D6S1699* and 5.17 cR proximal to *D6S1010*, clearly within the *TNDM* critical region.

Physiological role of *ZAC/PLAGL1*

A number of functional properties of *ZAC/PLAGL1* have been demonstrated recently (17,18,27–29,31). *ZAC*, *PLAGL1* and *LOT1* share coding sequences with a zinc finger motif (27–29) but have different 5′ ends (see above). Using differential display, *LOT1* was originally identified as a gene whose expression is lost on transformation in ovarian cancer, suggesting a tumor suppressor function. Such a function is supported by the fact that ZAC has similar activities to p53 in inducing cell cycle arrest and apoptosis, in LLC-PK1 kidney epithelial and SaOs-2 osteosarcoma cells (17).

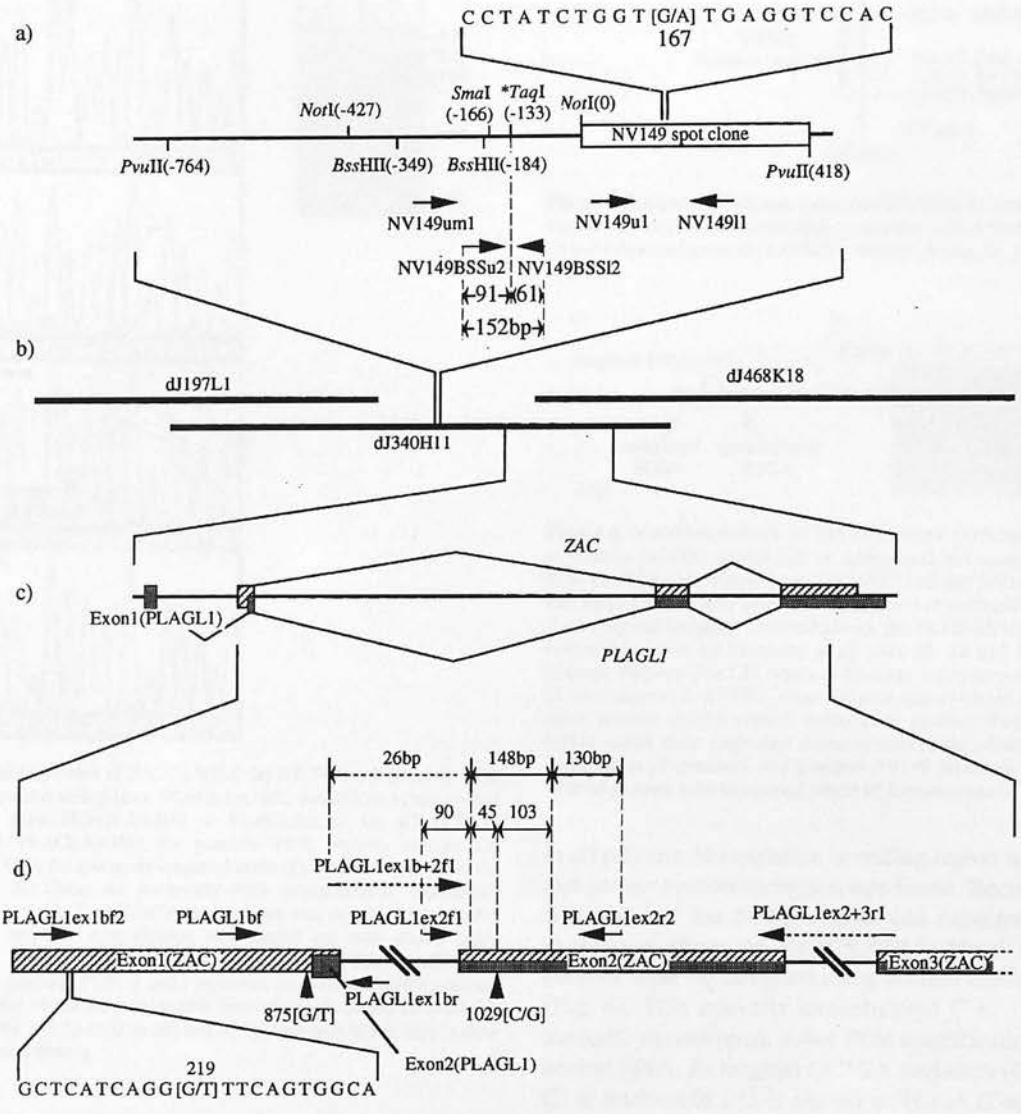


Figure 3. Restriction enzyme and contig map around *NV149*. (**a**) Restriction enzyme map for *Not*I, *Bss*HII, *Sma*I and *Pvu*II. The open box indicates the region covered by the spot clone. This clone was used for the Southern hybridization analysis in Figure 2. Numerals in parentheses refer to position relative to the *NV149 Not*I site. (**b**) PAC contig map. This is derived from the results of the chromosome 6 sequencing project (http://www.sanger.ac.uk/HGP/Chr6). (**c**) *ZAC/PLAGL1* gene structure. Three alternative forms of this gene, *ZAC*, *PLAGL1* and *LOT1*, were reported. *ZAC* consists of 3 exons (hatched area), named Exon1(ZAC), Exon2(ZAC) and Exon3(ZAC). The PLAGL1 transcript also consists of 3 exons (gray area), Exon1(PLAGL1), Exon2(PLAGL1) and Exon3(ZAC). Exon2(PLAGL1) lies within the 3′ part of Exon1(ZAC), sharing the same intron donor site. The acceptor site of Exon2(PLAGL1) is thus located inside Exon1(ZAC). *LOT1* is a chimeric clone (see text). (**d**) Exon structure and primers used in this study. Nucleotide positions refer to the *ZAC* cDNA sequence, AJ006354. Polymorphisms used for analysis of imprinted expression are indicated by arrowheads. Horizontal arrows represent primers. The sizes of fragments generated by *Ava*II digestion for analysis of the nucleotide 1029 polymorphism are indicated.

However, *ZAC* was originally identified through its ability to induce expression of the gene for the type 1 receptor for PACAP (17,18). Initially, *Zacl*, the mouse ortholog of *ZAC*, was isolated in an expression cloning experiment that also identified the $PACAP_1$-R itself and p53 (17). Importantly, it has been recognized for some time that PACAP is produced by pancreatic neural cells and is an extraordinarily potent autocrine regulator of glucose-stimulated insulin release (19,20). This immediately suggests that *ZAC* may be involved in the pathogenesis of TNDM, as a result of its effects on $PACAP_1$-R expression in the islet, and quite possibly also as a mediator of β-cell apoptosis (recalling that most TNDM patients eventually develop permanent diabetes).

More recently, Maebayashi *et al.* found that *ZAC/PLAGL1* is expressed at a very high level in the suprachiasmatic nucleus (SCN) during the early postnatal days (32). In mammals, the SCN is a central circadian pacemaker that regulates numerous behavioral and physiological rhythms. The similarity between the time course of the *ZAC/PLAGL1* expression and that of TNDM provide further support for the hypothesis that *ZAC/PLAGL1* is responsible for TNDM.

Genetic analysis of TNDM and its causative mechanism

In TNDM patients, several genetic abnormalities have been described: paternal uniparental disomy (UPD) of 6q24, segmental

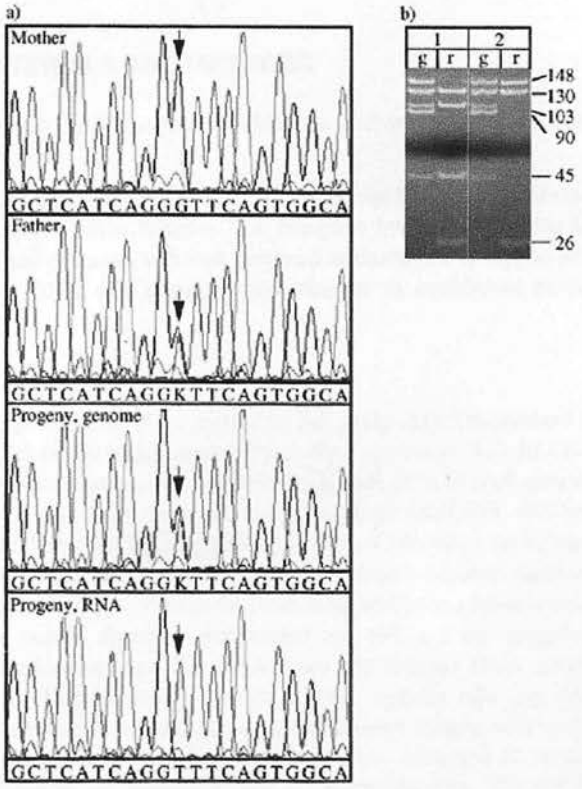


Figure 4. Imprinted expression of *ZAC/PLAGL1*. (**a**) RT–PCR and genomic PCR products were sequenced with primer, PLAGL1ex1bf2. Amplification was carried out with primer pairs PLAGL1ex1bf2 + PLAGL1ex2r2 for RT–PCR or PLAGL1ex1bf2 + PLAGL1ex1br1 for genomic PCR. Arrows indicate the polymorphic base. Only the paternally inherited allele (T) could be detected in the RT–PCR product. (**b**) Using the nucleotide 1029 polymorphism, monoallelic expression of mixed *PLAGL1*- and *ZAC*-type transcripts was examined in placentae (lanes 1g, 1r, 2g and 2r). Amplification was carried out with primer pairs, PLAGL1ex1b+2f1 and PLAGL1ex2r2 for RT–PCR, or PLAGL1ex2f1 and PLAGL1ex2r2 for genomic PCR. g and r represent genomic and cDNA-derived products, respectively. When the polymorphic base was G, it created an *Ava*II site which fragmented the 148 bp band to 103 and 45 bp. Two individuals show a clear monoallelic expression pattern.

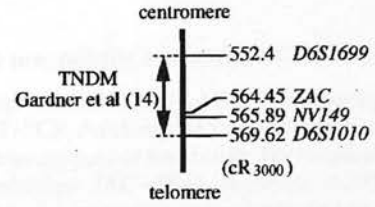


Figure 5. Radiation hybrid map around the *ZAC/PLAGL1* region. The Sanger Centre map constructed from the GeneBridge4 radiation hybrid panel is cited in this Figure (http://webace.sanger.ac.uk, SANGER_chrom6_rhmap_03_11_98).

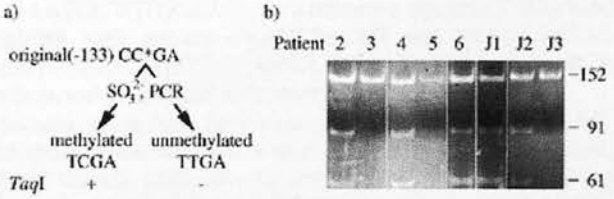


Figure 6. Methylation status of TNDM patients. (**a**) Schematic representation of sodium bisulfite conversion of methylated and unmethylated sequences. PCR amplification with primers NV149BSSu2 and NV149BSSl2 generates a 152 bp product. A new *Taq*I site is generated at nucleotide –133 (Fig. 3) only if the original template was methylated. (**b**) The DNA identification numbers follow the report by Hermann *et al.* (33). J1, J2 and J3 indicate Japanese patients. Patients 3 and J3, who are the cases with paternal uniparental disomy of chromosome 6 (UPD6), show only an unmethylated pattern, whereas the other patients exhibit normal methylation profiles. Both paternal alleles in UPD6 retain their imprinted demethylated status, whereas the differential methylation of maternal and paternal *NV149* alleles is also retained in the TNDM patients with biparental origin of chromosome 6.

trisomy of 6q24, dominant inheritance with a normal karyotype, as well as cases with no known genetic abnormality (14). The UPD6 cases suggest either a loss of function from a maternally expressed gene or increased expression of a paternally expressed one. The former possibility is excluded by the existence of patients with TNDM due to paternal duplication (partial trisomy) of 6q24. The cause of TNDM is therefore likely to be overexpression of a paternally expressed imprinted gene.

For TNDM cases with normal karyotypes, therefore, the following are possible pathogenetic mechanisms: (i) upregulation of the causative gene, possibly due to biallelic expression in place of monoallelic ('imprinting mutation'); (ii) a coding region mutation which activates protein function; (iii) submicroscopic reduplication of paternal 6q24 resulting in two expressed copies of the gene.

We examined the eight TNDM patients including two patients with uniparental paternal disomy (one Hungarian and the other Japanese) and six patients with a normal biparental chromosome complement. We first sequenced *ZAC/PLAGL1* in all patients. No mutation in coding region nor in the 5′-UTR and putative promoter region was found. Because the methylation status of the *NV149* region was expected to indicate the expression status of *ZAC/PLAGL1*, blood DNA from the patients were investigated using sodium bisulfite deamination (Fig. 6). This converts unmethylated C to U, but leaves 5-methylC unconverted. After PCR amplification of bisulphite-treated DNA, an original CC*GA sequence (C* is methylated C) at nucleotide 133 is altered to TCGA (*Taq*I site), whereas CCGA (unmethylated) is converted to TTGA. A 152 bp fragment was PCR-amplified from bisulphite-treated patient DNA and assayed with *Taq*I. In Figure 6, lanes 3 and J3, patients 3 (33) and J3 who are the Hungarian and Japanese paternal UPD6 cases, respectively, show, as expected, an unmethylated pattern of both alleles.

The methylation profiles of the TNDM patients with normal biparental DNA (Fig. 6, lanes 2, 4, 5, 6, J1 and J2) showed monoallelic methylation patterns. This indicates that maternal imprinting is not erased in these patients, and consequently lends no additional support to implicate *ZAC* in TNDM. Unfortunately, the expression of *ZAC/PLAGL1* in white blood cells is at very low level, and upregulation could not be reliably demonstrated. Definitive proof that *ZAC* is the TNDM gene may have to await the availability of other pathological material from TNDM patients, so that experiments can be performed to search for a switch from monoallelic to biallelic expression. None the less, the collective data strongly suggest *ZAC/PLAGL1* as a candidate gene for TNDM.

MATERIALS AND METHODS

Human parthenogenetic chimera and androgenetic mole samples

Blood samples were obtained from patient FD and age-matched, normal Scottish females. The complete hydatidiform mole and normal placental villi was prepared as described by Arima *et al.* (34). DNA was extracted according to an established method (35).

RLGS-M

RLGS was originally reported by our group (21). The protocol for RLGS in this experiment was described elsewhere (11). RLGS-M was performed as described by Kawai *et al.* (22). In brief, genomic DNA was blocked by nucleotide analogs (dGTPαS, dCTPαS, ddATP and ddTTP) with DNA polymerase I to reduce background. The DNA was digested with a methylation-sensitive restriction enzyme and then 5′-protuding ends were labeled in a fill-in reaction. The second digestion was carried out with a 6 bp recognizing restriction enzyme. The labeled and size-reduced DNA samples were electrophoresed through a 0.8% agarose tube gel (first-dimension separation). These samples were treated with a third restriction enzyme. In-gel samples were subjected to second-dimension 5% polyacrylamide gel electrophoresis. The gel was dried and autoradiographed. In this study, the restriction enzyme combinations *Asc*I–*Eco*RV–*Mbo*I and *Not*I–*Pst*I–*Pvu*II were used. The genomic DNA fragment corresponding to an RLGS-M spot was cloned according to a previously described protocol (11).

RH mapping

The GeneBridge 4 Radiation Hybrid Mapping Panel was purchased from Research Genetics (Huntsville, AL). Amplification was carried out with primers, NV149v2u1 (GCGAGGAGGGTGTGCCTTTG) and NV149v2l1 (GGCCCGTTGGCGAGGTTAGAG), under the following conditions: denaturation at 94°C for 2 min and 30 cycles of 96°C for 30 s, 60°C for 30 s and 72°C for 1 min. PCR products were electrophoresed in 3% agarose gel. The Sanger Centre 1998 GENE MAP SERVER for the GB4 panel (http://www.sanger.ac.uk/RHserver/RHserver.shtml) was used for calculation of the map position.

Methylation analysis of the *NV149 Not*I site

Genomic DNAs (10 µg) from placentae and peripheral blood lymphocytes were separated into two aliquots which were subsequently treated with or without methylation-sensitive enzymes *Not*I, *Bss*HII or *Sma*I. Both aliquots were then amplified with primers NV149v2um1 (ACGGCATCTGCCATTTGTCA) and NV149v2l1. PCR conditions were the same as for RH mapping. PCR products were sequenced using primer NV149v2u1. BigDye terminator reaction chemistry with an ABI 377 DNA sequencer was used. For Southern blotting, 5 µg of digested DNAs were treated with *Pvu*II and electrophoresed in a 1% agarose gel. According to standard protocols, DNAs were transferred to nylon membrane and were hybridized with a probe of *NV149* spot clone, labeled by random priming.

Imprinted or monoallelic expression of *ZAC/PLAGL1*

Monoallelic expression of *ZAC/PLAGL1*, was investigated in three ways. First, RT–PCR direct sequencing was carried out in a Japanese family, heterozygous at nucleotide 219 (nucleotide positions refer to the published *ZAC* cDNA sequence, AJ006354). RNA (3 µg) from placenta was reverse transcribed using oligo-d$(T)_{18}$ and SuperScriptII (Stratagene, La Jolla, CA) according to the manufacturer's protocol. As a PCR reaction template 1/50 of the reaction mixture was used with primers, PLAGL1ex1bf2 (AACGTTAATAAATCACTTAGGCGAGA) and PLAGL1ex2r2 (CTGACCAAATGCTGTGCCAT). As a reference, genomic DNAs were amplified with primers PLAGL1ex1bf2 and PLAGL1ex1br1 (ACCTCCAGCATGTTCTTGCC). RT–PCR and genomic PCR products were sequenced with primer PLAGL1ex1bf2.

Second, using three independent first-trimester fetuses (AC, HM and JC) heterozygous at nucleotide 875, six tissues (heart, kidney, muscle, cord, adrenal and lung) were analyzed. Using mismatched primer, ZAC-RT5′ (GGAATGTTTTCCTAGCTTCATTCCCGTAC) and downstream primer, ZAC-RT3′(2) (GGACCTCTCAGCTGTCACTAGCT), PCR on the RT product generated a 205 bp product. Digestion with *Bsi*WI produced bands of 205 bp if the polymorphic base was T (uncut), or 176 bp if the polymorphic base was G.

Third, PCR–RFLP was carried out using a polymorphism at nucleotide 1029. cDNAs from whole blood cells and placentae were subjected to PCR amplification with primers, PLAGL1ex1bf1 (TCAACCTTCTTGCACTGCAAA) and PLAGL1ex2+3r1 (ATGGGTAGCCATATGCCTCA). PCR conditions were 2 min at 94°C, followed by 25 cycles (96°C, 60°C and 72°C for 30 s, 30 s and 1 min, respectively). PCR products (1/20) were further amplified using nested primers, PLAGL1ex1b+2f1 (CCTGTCACTCAGTAGCCAA) and PLAGL1ex2r2, by 20 cycles under the same conditions. Reference genomic DNAs were amplified with primers, PLAGL1ex2f1 (TGATTCTGAAGCGGTCAGGG) and PLAGL1ex2r2. These PCR products were digested with *Ava*II and electrophoresed through 6% polyacrylamide gels. Digestion with *Ava*II gave bands of size 148 bp if the polymorphic base was C, and 103 and 45 bp if it was G.

Methylation status of *NV149* in TNDM patients

Unmethylated cytosine was converted to uracil by sodium bisulfite using a standard protocol (36). Briefly, 0.5 µg of *Pvu*II-treated DNA was incubated at 37°C for 10 min in 56 µl of 0.2 M NaOH. Thirty microliters of 10 mM hydroquinone and 520 µl of 3.6 M sodium bisulfite (final concentration 3.1 M) were added. These samples were incubated at 55°C for 16 h, followed by desalting with Geneclean (Bio101, Carlsbad, CA) and dissolved in 50 µl distilled water. After addition of 5.5 µl of 3 M NaOH and incubation at room temperature for 5 min, samples were ethanol precipitated and dissolved in 10 µl of TE (10 mM Tris–HCl, 0.1 mM EDTA pH 8). One microliter was used as PCR template for amplification with primers, NV149BSSu2 (GGGGTAGTYGTGTTTATAGTTTAGTA) and NV149BSSl2 (CRAACACCCAAACACCTACCCTA), under the following conditions: denaturation at 94°C for 2 min and 35 cycles of 96°C for 20 s, 60°C for 30 s and 72°C for 1 min. PCR products were digested with *Taq*I and electrophoresed on a 6% polyacrylamide gel. The original sequence CCGA 133 bp upstream from the *Not*I site was converted to UCGA, if the C in the CpG dinucleotide was

methylated, or UUGA, if it was unmethylated. The 152 bp PCR product is digested by *Taq*I to 91 and 61 bp products if derived from a methylated allele.

ACKNOWLEDGEMENTS

We thank Verne M. Chapman (since deceased) and Gary Chapman for their technical assistance and advice. This study has been supported by Special Coordination Funds and a Research Grant for the RIKEN Genome Exploration Research Project and CREST (Core Research for Evolutional Science and Technology) of Japan Science and Technology Corp. (JST), both of which are funded by the Science Technology Agency in Japanese Government (Y.H.). This work was also supported by a Grant-in-Aid for Scientific Research on Priority Areas and Human Genome Program, from the Ministry of Education, Science and Culture, and by a Grant-in-Aid for a Second Term Comprehensive 10-Year Strategy for Cancer Control from the Ministry of Health and Welfare (Y.H.). H.J. is an MRC PhD student.

REFERENCES

1. Surani, M.A., Barton, S.C. and Norris, M.L. (1984) Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature*, **308**, 548–550.
2. McGrath, J. and Solter, D. (1984) Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell*, **37**, 179–183.
3. Cattanach, B.M. and Kirk, M. (1985) Differential activity of maternally and paternally derived chromosome regions in mice. *Nature*, **315**, 496–498.
4. Beechey, C.V., Cattanach, B.M. and Selley, R.L. (1999) Mouse imprinting data and references. MRC Mammalian Genetics Unit, Harwell, Oxfordshire, UK, World Wide Web Site (http://www.mgu.har.mrc.ac.uk/imprinting/implink.html).
5. Morison, I.M. and Reeve, A.E. (1998) A catalogue of imprinted genes and parent-of-origin effects in humans and animals. *Hum. Mol. Genet.*, **7**, 1599–1609.
6. Kelsey, G. and Reik, W. (1998) Analysis and identification of imprinted genes. *Methods*, **14**, 211–234.
7. Hayashizaki, Y., Shibata, H., Hirotsune, S., Sugino, H., Okazaki, Y., Sasaki, N., Hirose, K., Imoto, H., Okuizumi, H., Muramatsu, M. *et al.* (1994) Identification of an imprinted U2af binding protein related sequence on mouse chromosome 11 using the RLGS method. *Nature Genet.*, **6**, 33–40.
8. Pearsall, R.S., Shibata, H., Brozowska, A., Yoshino, K., Okuda, K., deJong, P.J., Plass, C., Chapman, V.M., Hayashizaki, Y. and Held, W.A. (1996) Absence of imprinting in U2AFBPL, a human homologue of the imprinted mouse gene U2afbp-rs. *Biochem. Biophys. Res. Commun.*, **222**, 171–177.
9. Strain, L., Warner, J.P., Johnston, T. and Bonthron, D.T. (1995) A human parthenogenetic chimaera. *Nature Genet.*, **11**, 164–169.
10. Kajii, T. and Ohama, K. (1977) Androgenetic origin of hydatidiform mole. *Nature*, **268**, 633–634.
11. Hayashizaki, Y. and Watanabe, S. (eds) (1997) *Restriction Landmark Genomic Scanning (RLGS)*. Springer-Verlag, Tokyo, Japan.
12. Temple, I.K., Gardner, R.J., Robinson, D.O., Kibirige, M.S., Ferguson, A.W., Baum, J.D., Barber, J.C., James, R.S. and Shield, J.P (1996) Further evidence for an imprinted gene for neonatal diabetes localised to chromosome 6q22–q23. *Hum. Mol. Genet.*, **5**, 1117–1121.
13. Abramowicz, M.J., Andrien, M., Dupont, E., Dorchy, H., Parma, J., Duprez, L., Ledley, F.D., Courtens, W. and Vamos, E. (1994) Isodisomy of chromosome 6 in a newborn with methylmalonic acidemia and agenesis of pancreatic beta cells causing diabetes mellitus. *J. Clin. Invest.*, **94**, 418–421.
14. Gardner, R.J., Mungall, A.J., Dunham, I., Barber, J.C., Shield, J.P., Temple, I.K. and Robinson, D.O. (1999) Localisation of a gene for transient neonatal diabetes mellitus to an 18.72 cR3000 (approximately 5.4 Mb) interval on chromosome 6q. *J. Med. Genet.*, **36**, 192–196.
15. Plass, C., Shibata, H., Kalcheva, I., Mullins, L., Kotelevtseva, N., Mullins, J., Kato, R., Sasaki, H., Hirotsune, S., Okazaki, Y. *et al.* (1996) Identification of Grf1 on mouse chromosome 9 as an imprinted gene by RLGS-M. *Nature Genet.*, **14**, 106–109.
16. Hayward, B.E., Kamiya, M., Strain, L., Moran, V., Campbell, R., Hayashizaki, Y. and Bonthron, D.T. (1998) The human *GNAS1* gene is imprinted and encodes distinct paternally and biallelically expressed G proteins. *Proc. Natl Acad. Sci. USA*, **95**, 10038–10043.
17. Spengler, D., Villalba, M., Hoffmann, A., Pantaloni, C., Houssami, S., Bockaert, J. and Journot, L. (1997) Regulation of apoptosis and cell cycle arrest by Zac1, a novel zinc finger protein expressed in the pituitary gland and the brain. *EMBO J.*, **16**, 2814–2825.
18. Ciani, E., Hoffmann, A., Schmidt, P., Journot, L. and Spengler, D. (1999) Induction of the PAC1-R (PACAP-type I receptor) gene by p53 and Zac. *Brain. Res. Mol. Brain. Res.*, **69**, 290–294.
19. Yada, T., Sakurada, M., Nakata, M., Shioda, S., Yaekura, K. and Kikuchi, M. (1998) Autocrine action of PACAP in islets augments glucose-induced insulin secretion. *Ann. N. Y. Acad. Sci.*, **865**, 451–457.
20. Filipsson, K., Sundler, F. and Ahren, B. (1999) PACAP is an islet neuropeptide which contributes to glucose-stimulated insulin secretion. *Biochem. Biophys. Res. Commun.*, **256**, 664–667.
21. Hatada, I., Hayashizaki, Y., Hirotsune, S., Komatsubara, H. and Mukai, T. (1991) A genomic scanning method for higher organisms using restriction sites as landmarks. *Proc. Natl Acad. Sci. USA*, **88**, 9523–9527.
22. Kawai, J., Hirotsune, S., Hirose, K., Fushiki, S., Watanabe, S. and Hayashizaki, Y. (1993) Methylation profiles of genomic DNA of mouse developmental brain detected by restriction landmark genomic scanning (RLGS) method. *Nucleic Acids Res.*, **21**, 5604–5608.
23. Hayashizaki, Y., Hirotsune, S., Okazaki, Y., Shibata, H., Akasako, A., Muramatsu, M., Kawai, J., Hirasawa, T., Watanabe, S. *et al.* (1994) A genetic linkage map of the mouse using restriction landmark genomic scanning (RLGS). *Genetics*, **138**, 1207–1238.
24. Bird, A.P. (1986) CpG-rich islands and the function of DNA methylation. *Nature*, **321**, 209–213.
25. Wake, N., Takagi, N. and Sasaki, M. (1978) Androgenesis as a cause of hydatidiform mole. *J. Natl Cancer Inst.*, **60**, 51–57.
26. Shibata H., Hirotsune, S., Okazaki, Y., Komatsubara, H., Muramatsu, M., Takagi, N., Ueda, T., Shiroishi, T., Moriwaki, K., Katsuki, M. *et al.* (1994) Genetic mapping and systematic screening of mouse endogenously imprinted loci detected with restriction landmark genome scanning method (RLGS). *Mamm. Genome*, **5**, 797–800.
27. Varrault, A., Ciani, E., Apiou, F., Bilanges, B., Hoffmann, A., Pantaloni, C., Bockaert, J., Spengler, D. and Journot, L. (1998) hZAC encodes a zinc finger protein with antiproliferative properties and maps to a chromosomal region frequently lost in cancer. *Proc. Natl Acad. Sci. USA*, **95**, 8835–8840.
28. Kas, K., Voz, M.L., Hensen, K., Meyen, E. and Van de Ven, W.J. (1998) Transcriptional activation capacity of the novel PLAG family of zinc finger proteins. *J. Biol. Chem.*, **273**, 23026–2332.
29. Abdollahi, A., Roberts, D., Godwin, A.K., Schultz, D.C., Sonoda, G., Testa, J.R. and Hamilton, T.C. (1997) Identification of a zinc-finger gene at 6q25: a chromosomal region implicated in development of many solid tumors. *Oncogene*, **14**, 1973–1979.
30. Kas, K., Voz, M.L., Roijer, E., Astrom, A.K., Meyen, E., Stenman, G. and Van de Ven, W.J. (1997) Promoter swapping between the genes for a novel zinc finger protein and beta-catenin in pleiomorphic adenomas with t(3;8)(p21;q12) translocations. *Nature Genet.*, **15**, 170–174.
31. Pagotto, U., Arzberger, T., Ciani, E., Lezoualc'h, F., Pilon, C., Journot, L., Spengler, D. and Stalla, G.K. (1999) Inhibition of *Zac1*, a new gene differentially expressed in the anterior pituitary, increases cell proliferation. *Endocrinology*, **140**, 987–996.
32. Maebayashi, Y., Shigeyoshi, Y., Takumi, T. and Okamura, H. (1999) A putative transcription factor with seven zinc-finger motifs identified in the developing suprachiasmatic nucleus by the differential display PCR method. *J. Neurosci.*, **19**, 10176–10183.
33. Hermann, R., Laine, A.P., Johansson, C., Niederland, T., Tokarska, L., Dziatkowiak, H., Ilonen, J. and Soltész, G. (2000) Transient but not permanent neonatal diabetes mellitus is associated with paternal uniparental isodisomy of chromosome 6. *Pediatrics*, in press.
34. Arima, T., Matsuda, T., Takagi, N. and Wake, N. (1997) Association of IGF2 and H19 imprinting with choriocarcinoma development. *Cancer Genet. Cytogenet.*, **93**, 39–47.
35. Sambrook, J., Fritsch, E. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
36. Clark, S.J., Harrison, J., Paul, C.L. and Frommer, M. (1994) High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.*, **22**, 2990–2997.